Central Asian Scientific Journal

VOL 4(32)
2026



ASTANA

ISSN: 3135-3061

Электронный научный журнал «Central Asian Scientific Journal»

Central Asian Scientific Journal

выпуск №4 (32), октябрь 2026 г.
Том 1
Основан в 2021 году (издается ежеквартально)

зарегистрирован в Комитете информации Министерства информации и общественного развития Республики Казахстан №KZ77VPY00147053 от 17.04.2026 г.

Taqyryptyq baǵyt:

- Pedagogıkalyq, qoǵamdyq-áleýmettik, tehnıkalyq, ekonomıkalyq jáne zań ǵylymdary
- Aqparattyq-komýnıkasıalyq tehnologıalar
- Teorıalyq jáne ǵylymı-praktıkalyq ǵylymı zertteýler

Тематическая направленность:

- Педагогические, общественно-социальные, технические, экономические и юридические науки
- Информационно-коммуникационные технологии
- Теоретические и научно-практические научные исследования

Thematic focus:

- Pedagogical, socio-political, technical, economic, and legal sciences
- Information and communication technologies
- Theoretical and scientific-practical research

Jarıalanatyn aqparattyń, dáıeksózderdiń jáne ózge de baıandamalardyń durystyǵy úshin avtor jaýapty bolady

За достоверность публикуемой информации, цитат и иных изложений ответственность несет автор

The author is responsible for the accuracy of the published information, quotes, and other statements.

ISSN: 3135-3061

© Central Asian Scientific Journal, 2026

"Central Asian Scientific Journal" elektrondyq ǵylymı jýrnaly aqparattyq agenttigi

№3 (31), 2026 j
Shyǵarý jıiligi –jylyna 4 nómir
2021 j. bastap shyǵady

Bas redaktor:
Baıdildınov T. J. – p. ǵ. k., professor

Redaksıalyq alqa:
Latypov R.H. - t. ǵ. d., prof., Qazan, Reseı
Radwan Labban – Plymouth College, United Kingdom
Safarov G.A. - e. ǵ. d., prof., Tashkent, Ózbekstan
Mýkasheva A.A. - z.ǵ. d., prof., L.N. Gýmılev atyndaǵy EÝU
Baıǵojanova D.S. - p. ǵ. k., HAA akademıgi
Kojasheva G.O. - p.ǵ. k.,docent, Abay atyndaǵy KazPÝU
Teleýev G.B. - PhD, QAÝ
Özdenbaev J.Ş.- t. ǵ. k., I.Jansügırov atyndaǵy JU
Nūrǧalieva S.A. – PhD, asistent professor, AITU

Qazaqstan Respýblıkasy Aqparat jáne qoǵamdyq damý mınıstrliginiń 17.04.2026 j. №KZ77VPY00147053 aqparat komıtetinde tirkelgen.

JK DOC, 010000, Qazaqstan Respýblıkasy, Astana q.

Информационное агентство Электронный научный журнал «Central Asian Scientific Journal»

№3 (31), 2026 г.
Периодичность – 4 номера в год
Выходит с 2021 года

Главный редактор:
Байдильдинов Т.Ж. – к.п.н., профессор

Редакционная коллегия:
Латыпов Р.Х. – д.т.н., проф., Казань, Россия
Radwan Labban – Plymouth College, United Kingdom
Сафаров Г.А. –д.э.н., проф., Ташкент, Узбекистан
Мукашева А.А. – д.ю.н., проф., ЕНУ им. Л.Н. Гумилева
Байгожанова Д.С. – к.п.н, академик МАИН
Кожашева Г.О. – к.п.н, доцент, КазНПУ им. Абая
Телеуев Г.Б. – PhD, КАУ
Узденбаев Ж.Ш. – к.т.н., ст.преподаватель, ЖУ им. И.Жансугурова
Нургалиева С.А. – PhD, ассистент.проф., AITU

Зарегистрирован в Комитете информации Министерства информации и общественного развития Республики Казахстан №KZ77VPY00147053 от 17.04.2026

ИП DOC, 010000, Республика Казахстан, г. Астана

Information Agency Electronic scientific Journal "Central Asian Scientific Journal"

No.3 (31), 2026
Periodicity: 4 issues per year
Since 2021

Editor-in-Chief:
Baidildinov T.Zh. – Ph.D., Professor

Editorial Board:
Latypov R.H. – Doctor of Technical Sciences, Professor, Kazan, Russia
Radwan Labban –Plymouth College, United Kingdom
Safarov G.A. – Doctor of Economic Sciences, Professor,, Tashkent, Uzbekistan
Mukasheva A.A. – Doctor of Law, Professor, L.N. Gumilyov ENU
Baigozhanova D.S. – Ph.D., Academician of the MAIN
Kozhasheva G.O. –c.p.s, Abay KazNPU
Teleuev G.B. – PhD, KAU
Uzdenbaev Zh.Sh. –Candidate of Technical Sciences, Zhansugurov ZhU
Nurgaliyeva S.A. – PhD, assistant prof., AITU

Registered with the Information Committee of the Ministry of Information and Public Development of the Republic of Kazakhstan No. KZ77VPY00147053 dated 17.04.2026.

IP DOC, 010000, Kazakhstan, Astana

© Central Asian Scientific Journal, 2026

МАЗМҰНЫ – СОДЕРЖАНИЕ – CONTENT

ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖӘНЕ ҚҰҚЫҚ ҒЫЛЫМДАРЫ - СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЕ И ЮРИДИЧЕСКИЕ НАУКИ - SOCIO-ECONOMIC AND LEGAL SCIENCES

Төлегенов Тимур Сейфуллаулы
ДИНАМИКА ТУРИСТСКИХ ПОТОКОВ И ЭФФЕКТИВНОСТЬ ИНФРАСТРУКТУРЫ ТУРКЕСТАНА: ОТ ЭКСТЕНСИВНОГО РОСТА К СМАРТ-УПРАВЛЕНИЮ 10

Alexander Gorevoy
WHICH INEQUALITY STATISTICS SURVIVE AN ADMINISTRATIVE BOUNDARY CHANGE? EVIDENCE FROM KAZAKHSTAN, 2010–2023 14

Rakhimov Djalil
DOES THE SHIFT TO CASHLESS PAYMENTS GO TOGETHER WITH FASTER GROWTH IN HOUSEHOLD CREDIT? EVIDENCE FROM THE REGIONS OF KAZAKHSTAN, 2023-2025 26

Kazhgerey Asan
HOW MANY FOOD BASKETS DOES AN AVERAGE WAGE BUY? WAGES AND FOOD PRICES IN THE 20 REGIONS OF KAZAKHSTAN, 2024-2026 37

Tagimova Alua
WHO WAS LEFT BEHIND? EDUCATION AND FINANCIAL INCLUSION IN KAZAKHSTAN AND THREE CENTRAL ASIAN NEIGHBOURS, 2021-2024 44

Wei Wei, Mynzhanova Gulzhakan, Niyetalina Gaukhar
DIGITAL–PHYSICAL INTEGRATION AND ECONOMIC PERFORMANCE IN INDUSTRIAL INTEGRATED ENERGY SYSTEMS: A SCENARIO-BASED ANALYSIS OF SMART GRID CONTROL, BATTERY ENERGY STORAGE, AND DIGITAL TWINS 51

Xu Fumin, Wei Wei, Ruziyeva Elvira, Moldabekova A.
FROM CULTURAL RESOURCES TO GLOBAL CONSUMER VALUE: THE BRIDGE–IMC MODEL AND THE INTERNATIONALIZATION OF BLACK MYTH: WUKONG UNDER THE BELT AND ROAD INITIATIVE 64

Ibraimkhan Alikhan
WHICH CRIMES CAN A 14-YEAR-OLD BE TRIED FOR? THE LIST IN ARTICLE 15 OF KAZAKHSTAN'S CRIMINAL CODE COMPARED WITH 30 OTHER JURISDICTIONS 80

Yerboluly Madi
DID IMPORTED CARS REALLY GET CHEAPER? CHINESE CARS AND THE AVERAGE CUSTOMS VALUE OF PASSENGER CARS IMPORTED INTO KAZAKHSTAN, 2015-2025 91

Салимбаев Ербол Аманжолович
АНАЛИЗ ИНВЕСТИЦИОННОЙ АКТИВНОСТИ КАК ИНДИКАТОР ИЗМЕНЕНИЙ ПОТЕНЦИАЛЬНОГО СПРОСА НА КОНСАЛТИНГОВЫЕ УСЛУГИ 99

Rambergenov Ramil
REGIONAL NEW-FIRM FORMATION IN UZBEKISTAN AFTER THE 2017 REFORMS: PERSISTENCE IN SCALE, CONVERGENCE IN INTENSITY, AND RECOMPOSITION TOWARD TRADE 107

Abdesh Dinara
FROM A WOMEN'S LEAD TO A MEN'S LEAD: THE GENDER GAP IN DIGITAL PAYMENTS IN KAZAKHSTAN AND THREE CENTRAL ASIAN NEIGHBOURS, 2021-2024 116

Amangeldi Aidana
DID SMALL BUSINESS IN KAZAKHSTAN REALLY SHRINK IN 2026? THE NEW TAX CODE, REGIONAL TAX RATES AND THE FALL IN THE NUMBER OF INDIVIDUAL ENTREPRENEURS 124

Cheng Che
BUILD, BUY, OR PARTNER? ENTRY MODE SELECTION AND RISK GOVERNANCE FOR WIND POWER INVESTMENT IN KAZAKHSTAN **133**

Wang Liang
FROM CAPACITY EXPANSION TO SYSTEM VALUE: COORDINATED DEVELOPMENT OF RENEWABLE ENERGY, GRID FLEXIBILITY AND ENERGY STORAGE IN KAZAKHSTAN **146**

Chen Jun
CONSTRUCTION-PERIOD CREDIT RISK MANAGEMENT FOR LARGE INFRASTRUCTURE PROJECT FINANCE IN KAZAKHSTAN **159**

Alniyaz Danial
HOW MUCH FUEL DOES A HYBRID REALLY SAVE? A SAME-MODEL COMPARISON OF HYBRID AND GASOLINE CARS IN US FUEL ECONOMY **173**

ТЕХНИКАЛЫҚ ҒЫЛЫМДАР - ТЕХНИЧЕСКИЕ НАУКИ - TECHNICAL SCIENCE

Nurmakhan Almira
ADVANTAGES OF FIBER OPTIC COMMUNICATION OVER TRADITIONAL COPPER CABLES **180**

Kozhabai Bereke
SECURITY CHALLENGES IN IOT-BASED TELECOMMUNICATION NETWORKS **188**

Azamatuly Yerassyl
WHY AI-GENERATED CODE FAILS A BENCHMARK: A FAILURE ANALYSIS OF THREE OPEN LANGUAGE MODELS ON 30 HUMANEVAL+ TASKS **197**

Nurzhan Sanzhar
A FEW SHELLS, A FEW OPERATORS: WHERE LOW EARTH ORBIT IS CROWDED IN AN ALTITUDE-INCLINATION MAP OF 16,620 ACTIVE SATELLITES **206**

Askarbek Nurislam
CAN THE WEATHER TELL WHICH DAYS IN ASTANA WILL BE POLLUTED? CLASSIFYING PM2.5 EXCEEDANCE DAYS WITH MACHINE LEARNING AND CHECKING THE DATA AGAINST A GROUND MONITOR **212**

Iliyas Arman
SHOOTING ON THE MOVE: HOW MUCH AIM A ROBOT LOSES BY IGNORING ITS OWN VELOCITY, AND WHETHER A VIRTUAL-TARGET CORRECTION RECOVERS IT **223**

Ануар Әлімжан Бахытжанұлы, Сексенбаева Айша Курманбековна
СРАВНИТЕЛЬНЫЙ АНАЛИЗ СТАНДАРТОВ И МЕТОДИК ОБУЧЕНИЯ КИБЕРБЕЗОПАСНОСТИ: NICE FRAMEWORK, ECSF И ПРАКТИКО- ОРИЕНТИРОВАННЫЕ ТЕХНОЛОГИИ **229**

Mukhamedzhan Yerganat
GROWTH OF GITHUB DEVELOPER COMMUNITIES IN KAZAKHSTAN, UZBEKISTAN AND KYRGYZSTAN, 2020-2026: EVIDENCE FROM THE GITHUB INNOVATION GRAPH **239**

ПӘНДЕР АРАЛЫҚ ҒЫЛЫМДАР – МЕЖДИСЦИПЛИНАРНЫЕ НАУКИ – INTERDISCIPLINARY SCIENCES

Albiyev Asser
HOW MUCH OF AN ARTICLE COUNT IS HUMAN WORK? DECOMPOSING THE CREATION OF FIVE WIKIPEDIAS, 2010–2025 ..... **251**

Amangeldy Nurbolat
HOW MUCH OF KAZAKHSTAN'S URBAN BUILT STOCK PREDATES ITS SEISMIC CODES, AND WHERE DOES THAT STOCK SIT RELATIVE TO HAZARD? **261**

Torekhan Sabrina
WHERE ALMATY WALKS ITS OWN WAY: INFORMAL SHORTCUTS IN OPENSTREETMAP AND THE DETOURS THEY AVOID **267**

ӘЛЕУМЕТТІК ЖӘНЕ ГУМАНИТАРЛЫҚ ҒЫЛЫМДАР - СОЦИАЛЬНО-ГУМАНИТАРНЫЕ НАУКИ - SOCIAL SCIENCES AND HUMANITIES

Мұқатаева Айзат Амантайқызы, Стамбекова Жанат Байғазықызы, Слямханова Гүлім Болатқызы, Қызырбек Аяулым Ерболқызы
ЖАСАНДЫ ИНТЕЛЛЕКТ ДӘУІРІНДЕГІ ҒЫЛЫМИ ТАНЫМ: ПОППЕР, КУН ЖӘНЕ ЛАКАТОС ИДЕЯЛАРЫ ТҰРҒЫСЫНАН ТАЛДАУ **272**

Мұқатаева Айзат Амантайқызы, Сапарханова Мадина Төлегенқызы, Оженова Аружан Муратхановна, Тұрсынғазыева Ақерке Ерболатқызы
ШӘКӘРІМ ФИЛОСОФИЯСЫНДАҒЫ ЕРКІНДІК ЖӘНЕ МОРАЛЬДЫҚ ТАҢДАУ МӘСЕЛЕСІ **282**

Komekova Saida
DO UNHAPPY REVIEWERS MIX IN MORE RUSSIAN? STAR RATINGS AND KAZAKH-RUSSIAN CODE-SWITCHING IN 175,002 ONLINE REVIEWS **291**

Zhumagali Yerbol
DOES A BIGGER MODEL CLOSE THE GAP? KAZAKH-LANGUAGE ACCURACY ON KAZAKHSTAN-SPECIFIC VERSUS UNIVERSAL SCHOOL SUBJECTS IN SMALL OPEN-WEIGHT LANGUAGE MODELS **304**

Abdrakhmanova Madina
DOES A MEDAL BUY ATTENTION? WIKIPEDIA VIEWS OF KAZAKHSTAN'S ATHLETES BEFORE AND AFTER THE PARIS 2024 OLYMPIC **308**

ПЕДАГОГИКАЛЫҚ ЖӘНЕ ПСИХОЛОГИЯЛЫҚ ҒЫЛЫМДАР – ПЕДАГОГИЧЕСКИЕ И ПСИХОЛОГИЧЕСКИЕ НАУКИ – PEDAGOGICAL AND PSYCHOLOGICAL SCIENCES

Kaldarbek Mergen
FAMILY OR SCHOOL? SOCIOECONOMIC COMPOSITION AND MATHEMATICS ACHIEVEMENT IN KAZAKHSTAN, PISA 2022 AND 2025 ..... **316**

Zhenis Adilzhan
WHICH OLYMPIAD PROBLEMS DO STUDENTS TAKE TO THE BOARD? CHOICES AND DEFENCES IN THE 8TH-GRADE LEAGUE OF KARAGANDY MATHBATTLES 2026 ..... **323**

Базарбаева Айнур Бектуровна
ДЕТСКИЙ ФОЛЬКЛОР И ПОДВИЖНЫЕ ИГРЫ КАК ИНСТРУМЕНТЫ ФОРМИРОВАНИЯ ОСНОВ ПРАВОВОЙ КУЛЬТУРЫ У ДЕТЕЙ СТАРШЕГО ДОШКОЛЬНОГО ВОЗРАСТА **330**

БИОЛОГИЯ ҒЫЛЫМДАР - БИОЛОГИЧЕСКИЕ НАУКИ - BIOLOGICAL SCIENCES

Bazhenova Diana
GROWTH DYNAMICS OF STERLET (Acipenser ruthenus L.) REARED IN A RECIRCULATING AQUACULTURE SYSTEM **333**

ТУРИСТИЧЕСКИЕ И СПОРТИВНЫЕ НАУКИ - ТУРИСТІК ЖӘНЕ СПОРТ ҒЫЛЫМДАРЫ - TOURISM AND SPORTS SCIENCES

Төлеп Жазира Маратқызы
ГАСТРОНОМИЯЛЫҚ ФЕСТИВАЛЬДЕРДІҢ ТУРИСТІК ДЕСТИНАЦИЯ ТАРТЫМДЫЛЫҒЫН ҚАЛЫПТАСТЫРУДАҒЫ РӨЛІ МЕН ДАМУ ҮРДІСТЕРІ **340**

ЕСТЕСТВЕННЫЕ НАУКИ - ЖАРАТЫЛЫСТАНУ ҒЫЛЫМДАРЫ - NATURAL SCIENCES

Нургельдина Асель Ермековна, Серикбаева Жанерке Алимжановна
ШЕКАРАСЫЗ ФИЗИКА: ӘР ЕЛДЕГІ ФИЗИКАЛЫҚ БІЛІМДІ САЛЫСТЫРУ **347**

ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖӘНЕ ҚҰҚЫҚ ҒЫЛЫМДАРЫ - СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЕ И ЮРИДИЧЕСКИЕ НАУКИ - SOCIO-ECONOMIC AND LEGAL SCIENCES

УДК 338.487

Төлегенов Тимур Сейфуллаулы
Магистрант 2 курса
Образовательная программа «Менеджмент в туризме»
НАО «Международный университет туризма и гостеприимства»
(г. Туркестан, Казахстан)

ДИНАМИКА ТУРИСТСКИХ ПОТОКОВ И ЭФФЕКТИВНОСТЬ ИНФРАСТРУКТУРЫ ТУРКЕСТАНА: ОТ ЭКСТЕНСИВНОГО РОСТА К СМАРТ-УПРАВЛЕНИЮ

Аннотация: В статье исследуется динамика и структурные изменения туристских потоков в быстрорастущей культурно-познавательной дестинации на примере города Туркестан. На основе данных, официально предоставленных руководством многофункционального туристского комплекса «Karavansaray», проведен сравнительный макроанализ за 2024–2025 годы. Исследование базируется на оценке относительных показателей — средней загрузки средств размещения (Occupancy) и географической структуры прибытий, что позволило соблюсти принципы коммерческой конфиденциальности. Выявлено увеличение доли иностранных посетителей и устойчивый рост загрузки инфраструктурного ядра (с 38,9% до 47,7%). Обоснована необходимость внедрения дестинационных управляющих организаций (DMO) и технологий смарт-туризма для предотвращения овертуризма.

Ключевые слова: туристская дестинация, туристический поток, инфраструктура гостеприимства, Туркестан, комплекс Karavansaray, управление дестинацией (DMO), смарт-туризм, загрузка номерного фонда.

Развитие индустрии туризма в Республике Казахстан на современном этапе характеризуется уверенным переходом от формирования базовой инфраструктуры к необходимости комплексного и интеллектуального управления туристскими территориями. Особое, стратегически важное место в этом процессе занимает город Туркестан — признанная духовная и историческая столица страны. Завершение масштабных инфраструктурных проектов, включая строительство международного аэропорта, обновление городских транспортных артерий и, в первую очередь, запуск многофункционального туристского комплекса «Karavansaray», обеспечило городу статус одной из наиболее динамично развивающихся дестинаций всей Центральной Азии. Однако столь резкий рост популярности направления закономерно порождает классическую проблему территориального менеджмента: возникает дисбаланс между постоянно растущим туристским потоком и реальной пропускной способностью отдельных культурно-исторических объектов (в частности, мавзолея Ходжи Ахмеда Ясави и исторического центра).

В современной научной литературе эффективность туристской дестинации традиционно оценивается через призму комплексного взаимодействия ее ключевых элементов. Согласно классической концепции интегрированного туристского продукта

Д. Бухалиса, любая дестинация представляет собой сложную социально-экономическую систему, состоящую из достопримечательностей, транспортной доступности и средств размещения [1]. В свою очередь, Р. Батлер в своей фундаментальной теории жизненного цикла туристской территории (Tourism Area Life Cycle, TALC) утверждает, что неконтролируемый, стихийный рост туристических прибытий неизбежно приводит дестинацию к стадии стагнации и последующего упадка, если управляющими органами своевременно не внедряются эффективные механизмы пространственного и административного регулирования потоков [2]. Данные теоретические подходы сегодня дополняются актуальными трендами, задаваемыми Всемирной туристской организацией (UNWTO), которые категорично подчеркивают необходимость перехода территорий к модели «смарт-дестинаций», где управление потоками базируется на анализе больших данных (Big Data) и глубокой координации всех участников рынка через профильные организации по управлению дестинацией (DMO) [3, 4].

Целью данного исследования является сравнительный анализ структурных изменений туристского потока и эффективности загрузки ключевой туристской инфраструктуры Туркестана за 2024–2025 годы, а также разработка практических рекомендаций по оптимизации управления данной дестинацией в условиях растущего спроса.

Для достижения поставленной цели нами был применен комплекс научных методов: дескриптивный анализ (на основе агрегированных данных о туристских прибытиях), метод сравнительного статистического анализа (сопоставление операционных показателей 2024 и 2025 годов) и метод концептуального моделирования процессов управления. В качестве основной эмпирической базы исследования выступили статистические отчеты, официально предоставленные для проведения научного анализа администрацией многофункционального туристского комплекса «Karavansaray» (в том числе данные отеля Karavansaray Hotels Turkistan). Важно подчеркнуть, что с целью строгого соблюдения коммерческой тайны предприятия и принципов академической этики, в исследовании полностью исключены абсолютные финансовые показатели выручки (Total Revenue) и точные объемы реализованного номерного фонда. Весь анализ базируется исключительно на относительных метриках: процентной операционной загрузке (Occupancy) и процентном соотношении географических рынков-источников.

Анализ географической структуры туристского потока позволил выявить существенные сдвиги в рыночном позиционировании Туркестана. В 2024 году поток носил ярко выраженный внутренний, локальный характер. На долю граждан Республики Казахстан приходилось абсолютное большинство прибытий, тогда как международный сегмент был представлен преимущественно транзитными пассажирами или узкими целевыми паломническими группами из Турции, Узбекистана и Китая. Однако данные за 2025 год демонстрируют качественную трансформацию дестинации и значительный, статистически подтвержденный рост доли въездного туризма.

Согласно проанализированной статистике, предоставленной комплексом «Karavansaray», в 2025 году Туркестан значительно усилил свои позиции на международной арене туризма. Доля туристов из Турции выросла практически вдвое (составив более 11% от общего объема ключевых рынков), а доля китайских посетителей, проявляющих растущий интерес к историческому наследию Шелкового пути, увеличилась до 4,9%. Также наблюдается устойчивый и планомерный спрос со стороны стран Центральной Азии (в первую очередь Узбекистана и Кыргызстана). Подобная

интернационализация туристического потока свидетельствует о том, что Туркестан плавно переходит из статуса исключительно локальной паломнической святыни в категорию полноценного международного культурно-познавательного хаба.

Этот геоэкономический переход накладывает совершенно новые, повышенные требования на инфраструктуру гостеприимства всего региона. Рост числа иностранных посетителей требует немедленной адаптации городской сервисной среды: повсеместного внедрения многоязычной навигации, повышения международных стандартов обслуживания в сегменте общественного питания (Food & Beverage), а также диверсификации экскурсионных продуктов, которые должны быть понятны и интересны не только внутреннему туристу, обладающему базовым историческим контекстом, но и гостям из дальнего зарубежья.

Помимо глубоких структурных изменений потока, важнейшим макроэкономическим индикатором эффективности дестинации выступает степень операционной утилизации (загрузки) ее номерного фонда. Сравнительный анализ относительных показателей продемонстрировал уверенную и стабильную позитивную динамику. Если в 2024 году средняя годовая операционная загрузка (Occupancy) качественных средств размещения на примере анализируемого комплекса составляла 38,93%, то по итогам 2025 года этот показатель вырос до 47,71% (прирост составил существенные 8,78 процентных пункта). Данный факт эмпирически подтверждает, что построенная в городе масштабная инфраструктура (в первую очередь комплекс Karavansaray) не простаивает, а активно интегрируется в туристское потребление и успешно абсорбирует растущий рыночный спрос.

Тем не менее, детальный помесячный анализ выявляет наличие сильно выраженной фактора сезонности. Пиковые нагрузки традиционно и ожидаемо приходятся на весенние месяцы (март-апрель) в связи с празднованием государственного праздника Наурыз и наиболее благоприятными климатическими условиями региона, а также на осенний период (сентябрь-октябрь). Именно в эти периоды возникает серьезный риск кратковременного локального овертуризма, который на практике сопровождается образованием длительных очередей у ключевых достопримечательностей, пиковой нагрузкой на транспортную и санитарно-гигиеническую логистику, а также временным снижением общего качества предоставляемого сервиса.

Экстенсивный путь развития (подразумевающий рост исключительно за счет наращивания бюджетов на маркетинг и непрерывного строительства новых отелей) на данном этапе своего жизненного цикла исчерпывает себя. Дестинации Туркестана требуется качественный скачок и переход к интенсивному смарт-управлению. На основе проведенного теоретико-эмпирического анализа мы формулируем следующие научно-практические рекомендации, направленные на повышение общей эффективности туристской дестинации:

1. Институционализация координации через создание единого DMO. Туркестану критически необходима профессиональная локальная организация по управлению дестинацией (Destination Management Organization), которая объединит разрозненные усилия местного акимата, дирекции государственного историко-культурного заповедника-музея «Хазрет Султан», отельеров (включая менеджмент флагманского комплекса «Karavansaray»), ресторанного бизнеса и туроператоров. Создание такой структуры позволит синхронизировать городской календарь событий, унифицировать

маркетинговые послания на внешние рынки и применять единые, высокие стандарты качества обслуживания туристов на всей территории города.

2. Внедрение технологий смарт-мониторинга и предиктивной аналитики. Для эффективной борьбы с сезонными и событийными перегрузками инфраструктуры целесообразно использование систем геотрекинга и формирования тепловых карт плотности толпы на основе обезличенных данных мобильных операторов и систем городской сети Wi-Fi. Прямое информирование туристов в режиме реального времени о критической загруженности Мавзолея Ходжи Ахмеда Ясави через специализированные мобильные приложения (Smart City App) позволит заблаговременно перенаправлять туристический поток на альтернативные локации дестинации (многофункциональные развлекательные центры, новые интерактивные музеи, этно-аулы, гастрономические улицы).

3. Развитие стратегии пространственного и временного распределения маршрутов. Для решения проблемы «краткосрочных визитов» и увеличения продолжительности пребывания туристов (чтобы среднестатистический гость оставался на 2-3 ночи вместо быстрого транзитного визита на 1 световой день) необходимо формировать логически связанные многодневные маршруты. Такие туристические пакеты должны органично объединять центральное историческое ядро города Туркестан с мощными периферийными археологическими, культурными и природными локациями всего Туркестанского региона (включая городища Отрар и Сауран, природный потенциал Сайрам-Угамского государственного национального природного парка и другие объекты).

В заключение следует с уверенностью отметить, что туристская дестинация Туркестана демонстрирует исключительно высокие темпы развития и в настоящий момент успешно проходит критически важную стадию консолидации на внутреннем и растущем международном туристских рынках. Документально подтвержденное увеличение операционной загрузки номерного фонда и начавшаяся диверсификация географии посетителей являются сильными позитивными трендами. Однако для обеспечения долгосрочной, надежной устойчивости и, что самое главное, сохранения сакральной аутентичности бесценного культурного наследия, городу жизненно необходимо ускоренное внедрение современных цифровых инструментов управления потоками и глубокая, ежедневная координация действий всех стейкхолдеров индустрии гостеприимства.

Список литературы:

1. Buhalis D. Marketing the competitive destination of the future // Tourism Management. — 2000. — Vol. 21, No. 1. — P. 97–116.

2. Butler R.W. The concept of a tourist area cycle of evolution: Implications for management of resources // Canadian Geographer. — 1980. — Vol. 24, No. 1. — P. 5–12.

3. Management of Tourism Flows in Destinations. – Madrid: UN Tourism (UNWTO), 2024. – 112 p.

4. Ivars-Baidal J.A., Celdrán-Bernabeu M.A., Mazón J.N. Smart destinations and the evolution of ICTs: a new scenario for destination management? // Current Issues in Tourism. – 2022. – Vol. 25(10). – P. 1581-1600.

5. Об утверждении Концепции развития туристской отрасли Республики Казахстан на 2023-2029 годы: Постановление Правительства Республики Казахстан от 28 марта 2023 года № 262. – Текст: электронный // Электронное правительство РК.

Central Asian Scientific Journal

UDC 332.1

Alexander Gorevoy

Grade 11 student
Tamos Education Linguistics School
(Almaty, Kazakhstan)

WHICH INEQUALITY STATISTICS SURVIVE AN ADMINISTRATIVE BOUNDARY CHANGE? EVIDENCE FROM KAZAKHSTAN, 2010–2023

Abstract: Kazakhstan changed its administrative map twice in five years: Shymkent separated from South Kazakhstan in 2018, and three new oblasts appeared in 2022. The published regional accounts mark neither change, so parent oblasts continue under unchanged names on smaller territory. Simulation studies disagree about which inequality measures survive such redistricting. This study tests the question against a real, dated change and against every way two of the sixteen pre-2018 regions could have been joined. Official gross regional product and population for 2010 to 2023 gave two panels of output per head: regions as published, and the same economy on a constant pre-2018 geography of sixteen units. Five measures were computed on both panels and then on all 120 single-pair recombinations. Three sign results are arithmetic identities, so only magnitudes are empirical. Across the actual reforms the population-weighted coefficient of variation, the Theil index and the Gini shift by at most 2.05 per cent and the finding of convergence holds, while the richest-to-poorest ratio is overstated by 18 to 56 per cent. That stability is not a property of the measures: 41 recombinations move the weighted coefficient of variation and 49 the Gini by more than the real reform did. Distortion arises only where a recombination touches an extreme, and one region dominates: of the 105 recombinations not involving Atyrau, none reaches the real change. No inequality statistic can be assumed to survive a boundary change; which ones do depends on whether the divided units sit at an extreme, and that is checkable in advance.

Keywords: regional inequality; Kazakhstan; administrative boundaries; modifiable areal unit problem; sigma convergence; gross regional product

Introduction. Kazakhstan is one of the most spatially unequal economies in Central Asia. Oil-producing western oblasts and the two largest cities produce output per head many times that of the southern agricultural regions, and the size of that gap is a standing subject of policy argument. Recent work finds the divide widening: Rodríguez-Pose, Bartalucci and Rau report that rapid aggregate growth has been accompanied by a widening territorial divide, with wealth concentrating in Almaty and Astana while southern regions lag [1], and a companion study drawing on Kazakhstan's first well-being survey documents wide disparities both between and within regions [2]. Turganbayev, studying sigma and beta convergence from 1993 to 2014, reports divergence over the full period but convergence from 2006 onward [3]. Tleuberdinova and colleagues observe that no methodological consensus exists for measuring territorial disparity [4].

Every such measure is computed across whatever territorial units a statistical agency happens to publish, and those units are not fixed by nature. Geographers have long recognised this as the modifiable areal unit problem: the result of a spatial analysis depends on how the

territory is carved up [6]. Work on the United States using nested Theil decomposition shows that measured inequality is sensitive to the geographical scale chosen [7], the decomposition machinery for separating within-group from between-group components is well established [10, 12], and convergence and inequality have been examined jointly across the regions of the European Union [11]. The problem becomes acute when the partition changes during the study period. He, Chung, Bayrak and Wang address exactly this for China, using a multi-scale decomposition of Guangdong to show how administrative boundary changes affect measurement, and note that the question has rarely been studied [5]. Related Chinese work finds that administrative borders shape economic outcomes at the county level [8] and that inequality patterns in western provinces follow their own trajectory [9].

Two papers frame the narrower question of which measures can be trusted. Portnov and Felsenstein tested how sensitive common inequality measures are to changes in the ranking, size and number of regions, using simulated distributions and permutation tests, and concluded that only the population-weighted coefficient of variation, known as Williamson's index, and the population-weighted Gini can be considered sufficiently reliable for countries with a small number of regions of varying population size [13]. Gluschenko raises a related but distinct objection: that population-weighted indices are conceptually inconsistent as measures of regional inequality, because what they estimate is interpersonal inequality across the national population [14]. His argument concerns interpretation rather than stability under redistricting, and because this study reports both weighted and unweighted measures, its central result does not depend on which side of the weighting dispute the reader takes. Portnov and Felsenstein establish their conclusion by simulation; the present study tests the same question against a real, dated, documented change in a real economy, which is the natural complement to their design.

For Kazakhstan since 2018, the assumption that a region means the same thing across a study period is false, and no published study appears to have quantified the consequence. Two changes occurred. In 2018 Shymkent, the country's third-largest city, was granted the status of a city of national significance and separated from South Kazakhstan oblast, whose remainder was renamed Turkistan. On 8 June 2022 a presidential decree created Abai from part of East Kazakhstan, Jetisu from part of Almaty oblast, and Ulytau from part of Karaganda. First-order units rose from sixteen to seventeen and then to twenty. In the published accounts the new units simply begin and the parent oblasts simply continue, under the same names, with unbroken series, on smaller territory. Nothing in the data files marks the change. East Kazakhstan's nominal gross regional product falls 25.4 per cent between the first half of 2021 and the first half of 2022, across a year of high inflation; read at face value this is a severe recession. Most of it is the departure of Abai.

Splitting one region into two changes measured dispersion even when no economic activity has changed, and whether it raises or lowers the measure depends on whether the split separates rich from poor. This study quantifies that effect for Kazakhstan, asks which statistics it damages and which it leaves usable, and then asks whether the damage is attributable to these decrees at all. The objective is deliberately narrow: not a new account of Kazakh regional development, nor a claim that existing findings are wrong, but a measurement result of the kind that is cheap to establish once and costly to leave unexamined. The study covers 2010 to 2023 at oblast level and treats output per head only; it does not address within-oblast inequality and does not attempt to explain the level of Kazakh regional inequality.

Materials and Methods. The design is a paired comparison. The same economic data are measured twice, once on the territorial units as published and once on units held constant, and the difference between the two is the quantity of interest. Because both panels are built from identical underlying figures, any divergence arises from the partition rather than from the economy. A paired comparison establishes that a difference exists; the falsification arm described below is what addresses what the difference should be attributed to.

Two official series published by the Bureau of National Statistics of the Republic of Kazakhstan were retrieved on 31 July 2026 and re-retrieved on 20 September 2026 from the national accounts and demography dynamic-table archives: gross regional product, cumulative by quarter with annual totals, 2008 to 2025, in million tenge (element 5923), and population by region, annual, 2000 to 2025, in persons (element 6584). All numbers reported below are from the September vintage; the Bureau marks its 2021 gross regional product column as revised, and the 2021 values differ from the July vintage used in an earlier draft. Three properties of the output file required correction, none documented in the file itself. First, values are labelled in millions of tenge but stored scaled by 10: the national figure for January to September 2024 appears as 84,962,626,099,999,997,952, which on division gives about 85 trillion tenge, consistent with a full-year nominal gross domestic product near 140 trillion. Second, each observation is a year-to-date cumulative total rather than a quarterly flow, so annual values were taken as the January-to-December observation. Third, complete annual totals end in 2023, which fixes the study period at fourteen years.

Output per head was computed as gross regional product divided by population rather than taken from the published per-capita series, because that series omits 2010 to 2013 and 2021, and 2021 falls immediately before the second boundary change. The 2023 population totals used are those of element 6584: national 19,766,807, Almaty city 2,161,902, Astana 1,354,556, Atyrau 693,079, Turkistan 2,119,226 and Abai 610,198. South Kazakhstan oblast's series terminates at a January-to-March 2018 cumulative figure; because annual values are taken from the December observation, that partial-year figure has no annual counterpart and the unit simply leaves the annual panel after 2017. Turkistan and Shymkent both begin in 2018 and both carry a December 2018 observation.

The naive panel uses regions as published: sixteen units for 2010 to 2017, seventeen for 2018 to 2021, and twenty for 2022 to 2023. The harmonised panel maps every year onto a constant pre-2018 geography of sixteen units: Turkistan and Shymkent are recombined into South Kazakhstan, Abai into East Kazakhstan, Jetisu into Almaty oblast, Ulytau into Karaganda, and twelve unaffected units are carried unchanged. Output and population are summed, and output per head is recomputed as summed output over summed population. Five measures were computed on output per head for each year and panel: the coefficient of variation, unweighted and population-weighted; the population-weighted Theil index; the population-weighted Gini coefficient across regions; and the ratio of maximum to minimum.

Four comparisons were specified in a plan written before any result was computed. C1 is the distortion, naive minus harmonised, for each measure and year. C2 is the discontinuity in that distortion at each break. C3 regresses each measure on year by ordinary least squares with Newey–West standard errors at two lags, separately on each panel, testing whether the sign or significance of the trend differs. C4 was a placebo applying the same merge arithmetic to three never-redrawn oblast pairs in every year. Because fourteen observations is a small sample for an asymptotic standard error, the trend estimates were additionally re-examined with a moving-block bootstrap, block length three, 10,000 replications, fixed seed. A classification rule was

fixed beforehand: support required distortion of at least 5 per cent on at least two of the five measures at a break year together with a differing C3 conclusion. The plan also fixed a failure condition: if the placebo reproduced the break-year jump, the effect was to be reported as aggregation arithmetic and the paper was to become a methods caution rather than an inequality finding.

C4 as planned cannot fail on the extremes ratio, and this was discovered only after it had been run. It applies the same merges in every year, so it contains no event at a break year and structurally cannot produce a break-year discontinuity. C4 was therefore replaced by a timed sham split. For a pair of regions A and B, a sham history is constructed in which A and B are a single unit until a sham split year T and two separate units from T onward; the sham naive panel merges them before T and separates them from T, the sham harmonised panel merges them in every year. Before T the sham distortion is exactly zero; from T it is not, so the break-year jump is measured on the same footing as the real one. The decision rule was fixed in the analysis script before the sham block was run: if a sham 2018 split of never-redrawn regions reproduces a max/min jump of the same order as the real one, the failure condition has been met. The test was run for four named pairs at 2018 and 2022, then exhaustively for all 66 pairs drawn from the twelve never-redrawn units, and for all 120 pairs drawn from all sixteen. The plan is therefore best described as a pre-specified plan with one disclosed post-hoc operationalisation: the failure condition was fixed in advance, but the sham design that operationalises it was designed after the original placebo's results were known.

One property of the restricted pool has to be stated before its results are read. The region holding the minimum in thirteen of fourteen years, South Kazakhstan, is itself a redrawn unit, so it cannot appear in a pool defined as never-redrawn; that pool can only produce shams acting on the maximum. The exhaustive 120-pair pool is reported alongside it because it restores the minimum side, and results are broken out by which extreme a pairing touches. The study uses aggregate official statistics published for public use; no human subjects or private data are involved, and no ethical approval was required.

Results. Between 2010 and 2017 no boundary change had occurred, and the two panels return identical values on all five measures in all eight years. This is weaker than it looks: in those years the harmonisation performs no merge at all, so the check confirms that the loader and the measure code are consistent, not that the harmonisation logic is correct. The planned reconciliation was also run and passes: for every year, summed regional gross regional product equals the published national row to within floating-point noise, against a tolerance of 1 per cent. That reconciliation sums the published regional rows and compares them with the published national row; it never touches the recombination map, so it cannot detect an error in it. A mutation suite makes the point concretely: deliberately double-counting Abai, deleting Jetisu, or omitting Shymkent from the South Kazakhstan recombination each leave the reconciliation passing at 0.0000 per cent while the reported numbers move (the 2022 Gini distortion rises from 1.45 to 2.50 and 3.12 per cent respectively, and omitting Shymkent reduces the 55.89 per cent headline to zero). Those three mutations violate the partition or the mass-conservation checks and are caught by them. A fourth mutation, misfiling a successor unit under the wrong parent, violates neither check and is caught only by comparison with the decree text from which the crosswalk was built; the crosswalk's parentage therefore rests on the decrees, not on an automatic test.

Before any result is read, three properties of the measures must be stated, because they are true by arithmetic and none is a finding about Kazakhstan. Proposition 1 (extremes

compression): recombining two units replaces their two values by a single population-weighted average, which cannot exceed the larger nor fall below the smaller, so the harmonised ratio of maximum to minimum can never exceed the published one. Corollary: because the ratio depends on exactly two units, a recombination changes it if and only if one of the recombined units holds the maximum or the minimum, which is why 91 of the 120 pairings in Table 3 move it by exactly zero. Proposition 2 (the Theil gap is the within-group term): by the standard between-and-within decomposition of the Theil index [10, 12], the published index equals the harmonised index plus the population-weighted within-group inequality of the recombined units, so the published Theil is always at least the harmonised Theil; verified numerically, the 2018 difference equals the within-group term, 0.00764702, on both sides. Read this way, the Theil distortion is not noise: it is exactly the inequality inside the merged units that the pre-2018 boundaries concealed. Proposition 3: the largest distortion a minimum-side repartition can produce is fixed by the two lowest values, and the largest a maximum-side one can produce by the two highest; for 2018 these ceilings are 4.7711, 87.0974 and 96.0239 per cent, and the exhaustive search attains each exactly. What remains empirical is therefore the magnitude of the distortions the actual reforms produced, how often counterfactual repartitions exceed them, the behaviour of the trends, and which units drive the effect.

Table 1 – Inequality measured on published against harmonised regions, 2010–2023. Δ% is the distortion, naive minus harmonised, relative to the harmonised value.

Year	Units (naive / harm.)	CV unweighted Δ%	Gini Δ%	max/min naive → harm.	max/min Δ%
2010–2017	16 / 16	0.00	0.00	identical	0.00
2018	17 / 16	+1.83	+2.05	15.00 → 9.62	+55.89
2019	17 / 16	+2.15	+1.33	14.48 → 10.44	+38.71
2020	17 / 16	+2.25	+1.72	10.14 → 7.51	+35.04
2021	17 / 16	+2.03	+1.20	11.78 → 9.94	+18.43
2022	20 / 16	−0.57	+1.45	11.96 → 9.62	+24.42
2023	20 / 16	−1.67	+1.76	12.16 → 9.20	+32.14

Distribution-wide measures move little. The population-weighted Gini differs by at most 2.05 per cent and the unweighted coefficient of variation by at most 2.45 per cent, turning slightly negative in 2023. The ratio of richest to poorest region is overstated in every year after 2018, by between 18 and 56 per cent: for 2018 the published data imply that the richest region produced 15.0 times the output per head of the poorest, where constant geography gives 9.6. The sign of that column carries no information, as Proposition 1 states; only the magnitude is a result. The mechanism is visible in the arithmetic. Distribution-wide measures aggregate over sixteen or twenty units, so moving output between a parent and its offspring shifts them only at the margin. The extremes ratio depends on two units: separating Shymkent from South Kazakhstan left a lower-output rural remainder, Turkistan, at the bottom of the distribution, and

it is that fall in the minimum that drives the ratio. Atyrau holds the maximum in all fourteen years and is untouched by either reform (Figure 1).

Figure 1. The richest-to-poorest ratio on published vs constant geography

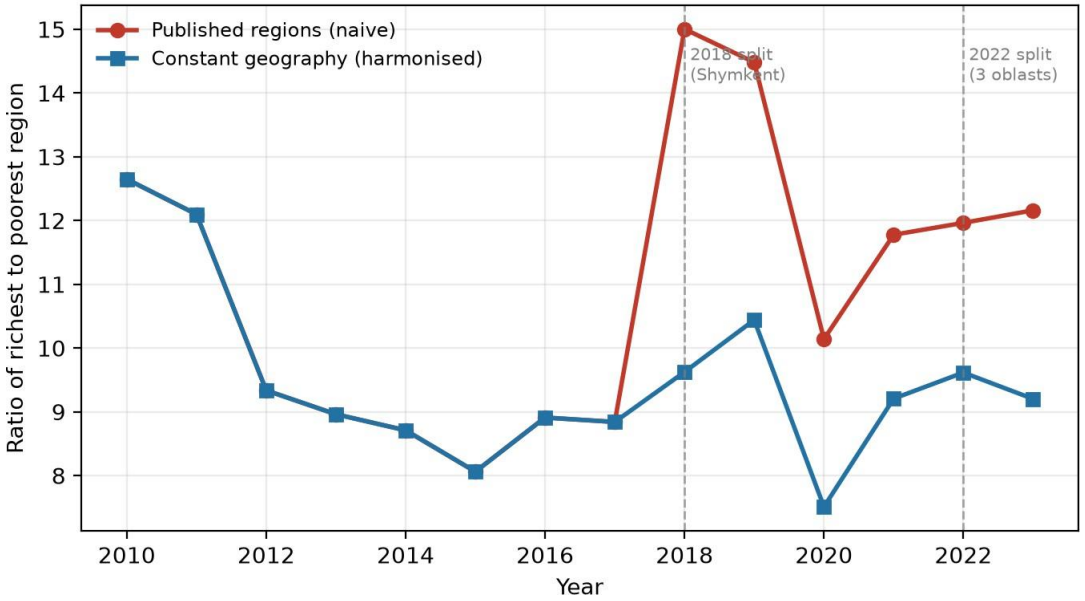


Figure 1 – The richest-to-poorest ratio on published against constant geography, 2010–2023. The two series coincide exactly before 2018.

Between 2017 and 2018 the distortion in the extremes ratio jumps by 55.89 percentage points, in the Theil index by 3.33 and in the Gini by 2.05. Between 2021 and 2022 the same jumps are +5.99, +0.53 and +0.25. The earlier change is far more consequential for measurement despite 2022 creating three new units against one in 2018, and the reason is not the number of units: Abai, Jetisu and Ulytau are middling regions and none of them, nor their parents, holds an extreme, whereas the 2018 change created a unit at the minimum.

Table 2 – Sigma convergence: ordinary least squares on year with Newey–West standard errors (two lags), n = 14, and moving-block bootstrap intervals on the residuals (block length 3, 10,000 replications, fixed seed).

Measure	Naive slope (p)	Harmonised slope (p)	Bootstrap 95% CI, naive	Bootstrap 95% CI, harmonised
CV unweighted	−0.00272 (0.229)	−0.00306 (0.210)	[−0.00664, +0.00169]	[−0.00677, +0.00101]
CV weighted	−0.00722 (<0.001)	−0.00754 (<0.001)	[−0.00995, −0.00426]	[−0.01034, −0.00459]
Theil	−0.00419 (<0.001)	−0.00465 (<0.001)	[−0.00579, −0.00249]	[−0.00630, −0.00297]
Gini	−0.00343 (<0.001)	−0.00399 (<0.001)	[−0.00485, −0.00196]	[−0.00549, −0.00250]
max/min	+0.15828 (0.242)	−0.13962 (0.205)	[−0.14754, +0.47637]	[−0.30770, +0.03641]

On the population-weighted coefficient of variation, the Theil index and the Gini, Kazakh regions converge across 2010 to 2023 on both panels, and harmonisation strengthens the estimated convergence slightly in each case, so the direction of Kazakh regional inequality does

not rest on the boundary artefact. The bootstrap intervals for the weighted coefficient of variation, the Theil index and the Gini exclude zero on both panels, so convergence on those measures is robust to serial correlation; the unweighted coefficient of variation is not distinguishable from zero on either panel. The extremes ratio behaves differently: its trend is positive on published units and negative on constant geography, but neither slope is distinguishable from zero and both bootstrap intervals are wide, so the reversal is inconclusive rather than established. All five measures are ratios and are unchanged by any deflator applied uniformly to regions within a year, which was verified numerically; region-specific price differences, which Kazakhstan does not publish at oblast level, are the one channel through which prices could enter.

The planned placebo, merging three never-redrawn pairs in every year, reads exactly 0.00 per cent on the extremes ratio in thirteen of fourteen years. Those zeros are an arithmetic necessity: neither Atyrau, which holds the maximum in all fourteen years, nor South Kazakhstan, which holds the minimum in thirteen, is in any placebo pair. In the single year where a pair member did touch an extreme, 2010, the placebo moved the ratio by −10.08 per cent, so a merge involving no boundary change at all can shift the ratio by double digits. The corrected timed sham splits are what the verdict rests on. A sham split of Atyrau from Mangistau at 2018, two regions that were never redrawn, distorts the extremes ratio by 38.85 percentage points, and the same sham at 2022 by 58.63, larger than the real 2018 figure of 55.89; a sham split of Almaty city from Almaty oblast, the closest structural twin of the Shymkent case, leaves the extremes ratio untouched while moving the Gini by 20.20 per cent. Across all 66 pairs of never-redrawn units the actual 2018 jump is unexceptional on every distribution-wide measure, sitting at or near the sham median with between 25 and 41 of the 66 repartitions moving those measures more, while on the extremes ratio the distribution is close to binary: 55 shams move it by exactly zero and the eleven that move it all involve Atyrau, seven of them by more than the real change.

Table 3 – All 120 sham pairings at split year 2018, classified by which extreme the pairing touches. The actual 2018 change gives +55.89 on max/min.

Pairing touches	Pairs	Pairs that move max/min	Magnitude range (pp)
The maximum (Atyrau)	15	15	38.85 – 96.02
The minimum (South Kazakhstan)	14	14	1.32 – 4.77
Neither extreme	91	0	exactly 0

The rule is exact rather than approximate: every one of the 29 pairings touching an extreme moves the ratio, every one of the 91 touching neither leaves it unchanged, and there are no intermediate cases. Magnitude is not symmetric: acting on the maximum moves the ratio by 38.85 to 96.02 percentage points, acting on the minimum by 1.32 to 4.77. The asymmetry matters because the 2018 decree acted on the minimum yet moved the ratio by 55.89 points, more than eleven times the largest minimum-side sham. A sham split returns two regions that already exist in the panel, so it can never produce a unit outside the observed range; the 2018 decree did. Turkistan's output per head in 2018 was 839,623 tenge against 1,308,920 for the undivided South Kazakhstan, 35.9 per cent below what had until then been the poorest region in the country. Under the decision rule fixed before the sham block was run, the failure condition

is nevertheless met: shams do reproduce and exceed the break-year jump. What the breakdown establishes is narrower and is reported as such: the reproduction comes entirely from repartitions acting on the maximum, and no repartition of existing regions acting on the minimum comes close.

Table 4 – Break-year distortion across all counterfactual single-pair repartitions, 2018. Percentages are absolute. The 2.05 per cent column asks how often a hypothetical repartition would have exceeded the largest distortion the actual reforms produced in any distribution-wide measure.

Measure	>2.05% (of 120)	>5% (of 120)	max % (120)	median % (120)	max % (66-pair)
CV unweighted	44	16	49.95	1.52	48.39
CV pop-weighted	41	24	50.16	0.78	32.31
Theil	60	42	96.81	2.05	47.21
Gini	49	24	38.44	1.23	17.53
max/min	28	15	96.02	0.00	87.10

Three things follow. The actual reform was gentle by comparison: between a third and a half of all single-pair recombinations of this economy would have moved the distribution-wide measures further than the real one did. The Theil index is the least reliable of the three population-weighted measures on every column. And the median tells the opposite story to the maximum for every measure: a typical repartition is nearly harmless, while an unlucky one is severe, which is why the question has to be asked case by case. The reference class rests on one region, and that must be said before Table 4 is used. Atyrau appears in 15 of the 120 recombinations, and removing it removes almost the entire effect: of the 105 recombinations not involving Atyrau, none moves the extremes ratio by as much as the real 2018 change, and the largest reaches 4.77 rather than 96.02. The three maxima of 50.16, 38.44 and 96.81 per cent are all produced by the same single recombination, Atyrau with South Kazakhstan; they are one fact about one pair of regions, not three independent results, and several of the largest values coincide because they saturate the closed-form ceilings of Proposition 3. Table 4 should therefore be read as a statement about how much influence a single extreme unit has in this economy, not as a survey of what redistricting does in general.

Two rules were fixed in advance and they point in different directions. Against the classification rule, only one of the five measures exceeds the 5 per cent threshold where the rule required two, although a C3 conclusion does differ; taken alone that would grade the result partially supported. Against the failure condition, the corrected placebo reproduces the break-year jump, and the plan states that in this event the paper becomes a methods caution rather than an inequality finding. The failure condition is the stricter rule and it governs; the claim that the effect is attributable to the boundary changes rather than to the act of merging units is withdrawn. Two qualifications are recorded rather than used to soften that reading. The shams that reproduce the real jump all act on the maximum while the 2018 decree acted on the minimum, so the reproduction is of the magnitude, not of the mechanism. And the failure condition fires because of one region: had Atyrau been absent from this economy, or had the reference class been built only from regions a redistricting could plausibly join, the condition would not have fired. Both qualifications are stated so that the reader can weigh the narrow base

on which the verdict rests: one region, in one recombination, inside a class that is a single-pair slice of the partition space rather than a survey of it.

Discussion. Two undocumented territorial discontinuities in Kazakhstan's regional accounts leave distribution-wide inequality measures and the finding of convergence substantially intact, and inflate the ratio of richest to poorest region by a quarter to a half. Neither effect is a property of the decrees. Both are properties of repartitioning, and their size is governed by which units the repartition touches. For extremes-based measures the governing rule is sharp: a repartition moves the ratio if and only if it creates or dissolves a unit at the maximum or the minimum, and otherwise moves it by exactly zero. For distribution-wide measures no such rule appears, and their apparent safety across the actual reforms does not generalise: every repartition moves them somewhat, and the real 2018 change happens to sit near the median. What the actual reforms demonstrate is that Kazakhstan divided units that happened to sit away from the tails, not that these statistics are safe to carry across a redistricting.

The practical implication is a diagnostic, and it is cheaper than the harmonisation it may render unnecessary. Before computing an inequality statistic across an administrative boundary change, an analyst should ask a single question: does any created, dissolved or shrunken unit sit at the maximum or minimum of the distribution? If none does, extremes-based measures are unaffected, not approximately but exactly, and in 91 of 120 Kazakh counterfactuals that is the answer. If one does, extremes-based measures are unusable without harmonisation, and the error can exceed 50 per cent. That question can be answered by inspection, before any panel is rebuilt. A second question follows for anyone who needs the size of the error: how unequal are the parts being separated, and which extreme do they sit at? Separating a rich city from a poor rural remainder at the bottom of the distribution, which is what 2018 did, moves the ratio far more than separating two ordinary neighbouring regions at the same end, and this second question needs the panel rebuilt. The diagnostic also explains the Kazakh case without recourse to anything Kazakh: 2018 looks dramatic and 2022 looks harmless, despite 2022 creating three times as many units, because 2018 created a unit at the minimum and 2022 did not.

This study does not reconcile with the existing literature as smoothly as an earlier draft claimed. Rodríguez-Pose and colleagues employ convergence analysis alongside two composite indices and conclude that Kazakhstan's territorial divide is widening [1]; the present study employs convergence analysis and finds that Kazakh regions converge on three of five measures, on both panels, across 2010 to 2023. These are the same method family reaching opposite conclusions, and harmonisation does not resolve the difference. One documented mechanism could account for part of it: Turganbayev reports divergence from 1993 to 2006 and convergence from 2006 onward [3], so a panel beginning in 2010 lies wholly inside the later, convergent regime. Whether that accounts for the difference cannot be determined here, because it would require matching study periods and measures directly. What this study adds is not a verdict on that dispute but a boundary on where the published units can be trusted, which is the same service He and colleagues performed for Guangdong [5] and which the modifiable areal unit literature predicts will be needed whenever a partition changes [6, 7].

On the narrower question of which measures withstand redistricting, the results qualify Portnov and Felsenstein rather than simply confirming them [13]. Across the actual reforms their two endorsed measures do hold: the population-weighted coefficient of variation and the Gini shift by at most 2.05 per cent, their convergence conclusions do not change, and the ratio of maximum to minimum moves by 24 to 56 per cent. Read alone, that corroborates them. It should not be read alone: the same 120-pairing exercise records what these measures do under

repartitions that did not happen, and there 41 recombinations move the weighted coefficient of variation and 49 the Gini by more than 2.05 per cent, with maxima of 50.16 and 38.44 per cent. The Theil index, which is not among the measures they endorse, is worse than both on every cut of the exhaustive pool: 60 of 120 repartitions exceed 2.05 per cent, 42 exceed 5 per cent, and its maximum is 96.81 per cent, which is a reason not to substitute it for them. What this adds to their result is therefore a boundary rather than an endorsement: their ranking survives, in that the extremes ratio fails far worse than anything else and their two measures are the best of the set, but it does not survive as a licence, because on real data a repartition can move even those two measures by tens of per cent. Gluschenko's objection to weighted indices [14] is about interpretation rather than stability and does not disturb the finding: the unweighted coefficient of variation, which is not subject to his criticism, is also displaced by at most 2.47 per cent, whereas the extremes ratio fails on either view, so readers on both sides of the weighting dispute are led to the same practical conclusion.

The objective was to quantify the distortion and to identify which statistics it damages. Both were met, and a third result arrived unbidden: the damage is predictable from the position of the affected units in the distribution, which makes it checkable in advance. The secondary hope, that harmonisation would overturn a published trend conclusion, was not met. Studies spanning an administrative boundary change should run the extreme-unit check before computing any extremes-based measure and should state which partition they use; where the check fails, units should be harmonised. Because roughly a third to a half of possible repartitions move the distribution-wide measures by more than Kazakhstan's real ones did, a stable Gini or coefficient of variation across a particular boundary change is evidence about that change, not a warrant for the next one. The check is worth running wherever administrative maps have changed during a study period, which is common across the post-Soviet region and beyond, and publishing agencies could remove the problem at source by flagging territorial discontinuities in the data files themselves.

Several limitations bound these results. The sham distribution is a calibration exercise over counterfactual partitions of one economy, not a sampling distribution; its exceedance proportions are not p-values. A sham split can never create a unit outside the observed range, whereas the 2018 decree placed Turkistan 35.9 per cent below the previously poorest region, so the sham distribution calibrates generic repartition arithmetic and does not bound the most extreme repartitions a government could enact. The reference class is a single-pair slice, a vanishing fraction of the roughly 10.5 billion ways sixteen units could be partitioned; it cannot contain the treatment it is compared against; and its members are not all enactable, since the recombination producing every quoted maximum joins Atyrau to South Kazakhstan, two oblasts roughly 1,800 km apart. The sham design was chosen after the original placebo's results were known, and a reader weighing the methods-caution scope should weigh that too; no sensitivity analysis over alternative sham designs was run, and the verdict rests on this one. The trend tests rest on fourteen annual observations with serially correlated errors; no formal stationarity or autocorrelation diagnostics beyond the Newey–West correction and the block bootstrap were run, and the regressions do not include a break dummy at 2018 or 2022. Units are harmonised backwards onto pre-2018 geography because a forward mapping would require sub-oblast output that the Bureau does not publish. Regional price differences are not observed, and the study treats output per head alone; inequality in well-being, services or opportunity may respond differently to the same boundary changes [2].

Deviations from the plan were recorded as they were made. The population source was changed before results, from the gapped per-capita series to population computed from element 6584. The planned placebo C4 was executed, found incapable of failing on the extremes ratio, retained with its limitation stated, and supplemented by the timed sham split, whose decision rule was fixed in the analysis script before the block was run once. The national reconciliation and the statement of the sign identities were specified in the plan but executed only in a later draft, and that lateness is recorded. The scope was changed by the failure condition from an inequality finding to a methods caution. And the sham results were broken out by extreme after an earlier draft had reported the never-redrawn pool as its headline distribution, a pool from which the minimum-holding region is excluded by construction.

Conclusion. Two administrative boundary changes that Kazakhstan's regional accounts do not mark inflate the published ratio of richest to poorest region by 24 to 56 per cent for every year from 2018, while the population-weighted coefficient of variation, the Theil index and the Gini move by at most 2.05 per cent and the finding of regional convergence survives on both partitions. The stability of the distribution-wide measures across these particular reforms is not a property of the measures: 41 of the 120 single-pair recombinations of the same economy move the weighted coefficient of variation, and 49 the Gini, by more than the real reform did, and the Theil index is the least stable of the three. What decides whether a statistic survives is whether the divided units sit at an extreme of the distribution, a condition that is exact for extremes-based measures and checkable by inspection before any panel is rebuilt. The falsification test fixed in advance was met, so the effect is reported as repartition arithmetic rather than as a property of these decrees, and that verdict rests on one region, Atyrau, in one recombination. A statistic can be arithmetically correct and still describe something that never happened: the twenty-five per cent fall in East Kazakhstan's output is real in the sense that the numbers say so, and imaginary in the sense that no factory closed. Keeping those two senses apart is most of what careful measurement means.

Acknowledgments. Input data are published by the Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan (dynamic tables, elements 5923 and 6584) and are available to any reader; every number reported here derives from those two series by the arithmetic described in the Methods. The author declares no conflict of interest and received no funding for this work.

References:

1. Rodríguez-Pose A., Bartalucci F., Rau G. One Kazakhstan, multiple nations: on a growing regional divide amidst economic dynamism. Asia-Pacific Journal of Regional Science, 2025, vol. 9, pp. 629–651. https://doi.org/10.1007/s41685-025-00385-7

2. Rodríguez-Pose A., Bartalucci F., Kurmanov B., Rau G., Nigmetov K. Assessing regional inequalities in Kazakhstan through well-being. Asian Development Review, 2024, vol. 41, pp. 301–333. https://doi.org/10.1142/S0116110524500033

3. Turganbayev Y. Regional convergence in Kazakhstan. Post-Communist Economies, 2016, vol. 28, pp. 314–334. https://doi.org/10.1080/14631377.2016.1204745

4. Tleuberdinova A., Nurlanova N., Alzhanova F., Tulenbayeva Z. Developing a methodology for measuring regional socio-economic development balance: a case study of Kazakhstan. Economies, 2026, vol. 14, p. 195. https://doi.org/10.3390/economies14060195

5. He S., Chung C. K. L., Bayrak M. M., Wang W. Administrative boundary changes and regional inequality in provincial China. Applied Spatial Analysis and Policy, 2018, vol. 11, pp. 103–120. https://doi.org/10.1007/s12061-016-9203-5

6. Pietrzak M. B. The modifiable areal unit problem – analysis of correlation and regression. Equilibrium, 2014, vol. 9, pp. 113–131. https://doi.org/10.12775/equil.2014.028

7. Lee A. S., Breau S. Scales of inequality: applying a two-stage Theil decomposition approach to examine the changing geography of income inequality. International Regional Science Review, 2026. https://doi.org/10.1177/01600176261440842

8. Zhang X., Li C., Li W., Song J., Yang C. Do administrative boundaries matter for uneven economic development? A case study of China's provincial border counties. Growth and Change, 2017, vol. 48, pp. 883–908. https://doi.org/10.1111/grow.12196

9. Qin X., Zou H., Wang L. Changing regional inequality patterns in western China: a case study of Xinjiang. Complexity, 2021, vol. 2021. https://doi.org/10.1155/2021/9160354

10. Akita T. Decomposing regional income inequality in China and Indonesia using two-stage nested Theil decomposition method. The Annals of Regional Science, 2003, vol. 37, pp. 55–77. https://doi.org/10.1007/s001680200107

11. Tselios V. Growth and convergence in income per capita and income inequality in the regions of the EU. Spatial Economic Analysis, 2009, vol. 4, pp. 343–370. https://doi.org/10.1080/17421770903114711

12. Frosini B. V. Approximation and decomposition of Gini, Pietra–Ricci and Theil inequality measures. Empirical Economics, 2012, vol. 43, pp. 175–197. https://doi.org/10.1007/s00181-011-0464-1

13. Portnov B. A., Felsenstein D. On the suitability of income inequality measures for regional analysis: some evidence from simulation analysis and bootstrapping tests. Socio-Economic Planning Sciences, 2010, vol. 44, pp. 212–219. https://doi.org/10.1016/j.seps.2010.04.002

14. Gluschenko K. Measuring regional inequality: to weight or not to weight? Spatial Economic Analysis, 2018, vol. 13, pp. 36–59. https://doi.org/10.1080/17421772.2017.1343491

UDC 336.7

Rakhimov Djalil

Grade 11 student

Supervisor: Kudaibergen Madi Askhatuly

Republican Physics and Mathematics School

(Astana, Kazakhstan)

DOES THE SHIFT TO CASHLESS PAYMENTS GO TOGETHER WITH FASTER GROWTH IN HOUSEHOLD CREDIT? EVIDENCE FROM THE REGIONS OF KAZAKHSTAN, 2023-2025

Abstract: Kazakhstan has moved from cash to card payments very quickly, and experiments in other countries show that people tend to spend more when they pay by card. This study asks whether the regions of Kazakhstan that moved away from cash faster also had faster growth in bank loans to households. I used National Bank of Kazakhstan data for all 20 regions: the value of cashless card payments and card cash withdrawals in 2023, 2024 and 2025, and outstanding household loans on 1 January 2024, 2025 and 2026. Across the 20 regions, growth in cashless payments in 2024 was strongly correlated with growth in household loans ($r = 0.66$, $p = 0.002$). A leave-one-out check showed that this result depended on one region, Zhetysu: without it, the correlation fell to 0.17 ($p = 0.48$). Zhetysu is one of three regions created in June 2022, and in 2024 these three regions had the fastest growth in card cash withdrawals in the country, while the regions they were separated from grew slowly or declined. When the new regions were excluded, or merged back into their former regions, none of the correlations was significant. The change in the cashless share of card turnover, which measures the move away from cash directly, was not related to consumer loan growth in either year. At the regional level, the shift to cashless payments has not come with faster borrowing, and regional statistics for 2023 and 2024 need care because of the 2022 boundary change.

Keywords: cashless payments; household credit; consumer loans; regional statistics; Kazakhstan; correlation; influential observations

1. Introduction

In Kazakhstan, paying by card or by phone has become the normal way to pay. According to the National Bank of Kazakhstan, cashless payments made with cards in the country reached 186.3 trillion tenge in 2025, compared with 142.6 trillion tenge in 2023, while the number of cash withdrawals with cards went down over the same period [1, 2]. The Global Findex survey shows the same change from the side of households: the share of adults in Kazakhstan who made or received a digital payment rose from 77.9% in 2021 to 85.0% in 2024 [3]. At the same time, bank lending to households grew fast. Outstanding bank loans to households in tenge increased from 16.7 trillion tenge on 1 January 2024 to 24.8 trillion tenge on 1 January 2026, and about two thirds of this amount are consumer loans [4].

There is a reason to think these two trends could be connected. Research in consumer psychology has shown for a long time that the way people pay changes how much they spend. Feinberg found that credit card cues alone made people more willing to spend [5]. In real

transactions of high value, Prelec and Simester found that people's willingness to pay could be up to twice as high when they were told to pay by credit card instead of cash [6]. Soman showed that spending depends on how strongly a payment method makes people notice the amount they pay and how soon the money actually leaves their account [7]. Raghubir and Srivastava found that people were willing to spend more when a credit card logo was present, and that they spent more with a stored-value certificate than with cash of the same value [8]. One explanation is the "pain of paying": giving up money is unpleasant, and this feeling is weaker when paying is separated from consuming [9]. Card payments have also been linked to more impulsive purchases [10].

Most of this evidence comes from experiments or from data on individual shoppers. At a larger scale, Humphrey described how cards replaced cash in consumer payments in the United States [11], and Li, Wu and Xiao found that the development of digital finance in China was associated with higher household consumption [12]. Borrowing on cards can also be a reasonable choice for a household when other kinds of credit are more costly to arrange [13].

Kazakhstan's regions give a natural setting to look at this question. The share of cashless payments in card turnover is very different from region to region: in 2023 it ranged from 69.4% in North Kazakhstan region to 91.4% in the city of Almaty [1]. If paying by card makes people spend and borrow more, then the regions that are moving away from cash faster might also be the regions where household debt grows faster. But this does not follow automatically. A pattern that holds for individuals does not have to appear in regional totals, and a correlation between regional figures can appear or disappear for reasons that have nothing to do with how individuals behave [14].

This study uses data for the 20 regions of Kazakhstan in 2024 and 2025 and asks three questions:

• Was faster growth in cashless card payments associated with faster growth in household loans?

• Was the move away from cash itself, measured as the change in the cashless share of card turnover, associated with faster growth in consumer loans?

• How much do the results depend on single regions, in particular on the three regions created in 2022?

2. Data and methods

Payments data. Data on payments come from the National Bank of Kazakhstan releases "Payment cards and electronic banking services". Table 5 of each release shows the use of payment cards in Kazakhstan by region: the number and value of cashless operations and the number and value of cash withdrawals for the previous calendar year. I used the releases of 29 January 2024 (No. 244), 30 January 2025 (No. 256) and 30 January 2026 (No. 268), which cover 2023, 2024 and 2025 [1, 2, 15]. In each year the 20 regions add up exactly to the national total in the table.

Loan data. Data on loans come from the National Bank files "Loans of the banking sector to the population by region" for 2024 and 2025 [4]. These files give outstanding loans at the start of every month for each region, split into national and foreign currency and, separately, into mortgage, consumer and other loans. In National Bank statistics, loans to the population are loans to individuals for purposes not related to business, issued by second-tier banks and the Development Bank of Kazakhstan [16]; loans that sole proprietors take for their business are

not included. I used balances in national currency on 1 January 2024, 2025 and 2026. Loans in foreign currency are too small to matter: they were 0.02% of all household loans on 1 January 2024 and 0.01% on 1 January 2026.

Regions. The analysis covers 17 regions and the cities of Astana, Almaty and Shymkent. Three of the regions are new. By Presidential Decree No. 887 of 3 May 2022, which came into force on 8 June 2022, Abai region was separated from East Kazakhstan region, Zhetysu region from Almaty region, and Ulytau region from Karaganda region. The administrative centre of Almaty region was moved from Taldykorgan, which became the centre of Zhetysu, to Qonaev [17]. This change matters for the results (Section 3.3).

Checking the data. Before the analysis, every value in the working spreadsheet was compared with the original publications: 240 payment values (20 regions, 3 years, 4 indicators) and 60 loan values (20 regions, 3 dates). No differences were found.

Measures. For each region and year I calculated:

• growth in cashless payments, the percentage change in the value of cashless card operations compared with the previous year;

• the cashless share, the value of cashless operations divided by the value of cashless operations plus cash withdrawals, in per cent, and its change in percentage points;

• growth in household loans and in consumer loans, the percentage change in outstanding loans from 1 January of one year to 1 January of the next.

Payments in a calendar year are compared with the change in loans over the same year. For 2024 this means cashless payments in 2024 compared with 2023, and loans on 1 January 2025 compared with 1 January 2024. The cash side of these data covers only cash that was taken out with a card. Cash that people receive in other ways and spend without a card is not in the data, so the cashless share describes card turnover, not all spending in a region.

Analysis. For each year I calculated the Pearson correlation coefficient r and the Spearman rank correlation ρ across the regions. With 20 observations, a correlation has to be at least 0.44 in absolute value to be significant at the 5% level in a two-sided test. Because one region is 5% of such a small sample, I also used a leave-one-out check: the correlation was calculated 20 more times, each time without one region. This is a simple form of the idea behind Cook's distance, which measures how much a result depends on one observation [18], and it follows Anscombe's advice to look at the data behind a correlation and not only at the coefficient [19]. Finally, the analysis was repeated in two ways that take the 2022 boundary change into account: first without the three new regions (17 regions), and then with each new region added back to the region it was separated from, which restores the boundaries before 2022 (17 units). The first calculations were made in Microsoft Excel with the CORREL function, and all results were then checked again in Python with the SciPy library.

3. Results

3.1 The national picture

Table 1 shows the main national figures. Cashless card payments grew by 15.9% in 2024 and by 12.8% in 2025. The value of cash withdrawals also grew, but more slowly, and the number of withdrawals went down every year, from 241.6 million in 2023 to 233.9 million in 2025. As a result, the cashless share of card turnover increased from 86.2% to 87.2%. Across

regions, the share was lowest in North Kazakhstan region (69.4% in 2023 and 72.9% in 2025) and highest in the city of Almaty in 2023 (91.4%) and in Almaty region in 2025 (91.7%).

Household loans grew by 23.9% in 2024 and by 19.8% in 2025. Consumer loans grew even faster in 2024, by 33.5%, and their share in all household loans rose from 61.8% to 67.3% between the start of 2024 and the start of 2026.

Table 1. Card payments and household loans in Kazakhstan, 2023-2025

Indicator	2023	2024	2025
Cashless card payments, trillion tenge	142.6	165.2	186.3
Card cash withdrawals, trillion tenge	22.9	25.4	27.3
Number of card cash withdrawals, million	241.6	238.4	233.9
Cashless share of card turnover, %	86.2	86.7	87.2
Household loans at the end of the year, trillion tenge	16.7	20.7	24.8
Consumer loans, % of household loans	61.8	66.6	67.3

Loans are balances in national currency on 1 January of the following year. Sources: National Bank of Kazakhstan releases No. 244, 256 and 268 (Table 5); regional loan files for 2024 and 2025.

3.2 Growth in cashless payments and growth in loans

In 2024, the regions where cashless payments grew faster also tended to have faster growth in household loans (Figure 1). The correlation was strong: $r = 0.66$ ($p = 0.002$) and $\rho = 0.54$ ($p = 0.015$). For consumer loans it was a little stronger, $r = 0.70$ ($p = 0.001$). In 2025 the relationship was much weaker: $r = 0.27$ ($p = 0.24$) for all household loans and $r = 0.27$ ($p = 0.26$) for consumer loans.

When I first got $r = 0.66$ for 2024, it looked like a clear answer to the main question. The leave-one-out check changed this (Figure 2). Removing any one of 19 regions left the correlation between 0.63 and 0.77. Removing Zhetysu region dropped it to 0.17 ($p = 0.48$), far below the significance threshold. For 2025, without Zhetysu, the correlation was -0.01. In other words, most of the 2024 result came from one point. In 2024 Zhetysu had by far the fastest growth in both variables: cashless payments rose by 63.5% (the median for all regions was 15.1%) and household loans by 49.7% (median 22.8%).

3.3 Zhetysu and the 2022 boundary change

Zhetysu's cashless share did not go up in 2024. It went down, from 82.6% to 81.7%, because card cash withdrawals in the region grew even faster than cashless payments, by 74.2%. So the region that produced the strong correlation was not moving away from cash at all.

Cash withdrawals in all regions (Figure 3) help to explain why. The three largest increases in 2024 were in exactly the three regions created in 2022: Zhetysu (+74.2%), Ulytau (+29.7%) and Abai (+29.5%). The regions they were separated from showed the opposite: Almaty region -32.2%, East Kazakhstan -6.2% and Karaganda +2.3%, while the median for the other 14 regions was +6.9%. Each of the three new regions had faster growth in cash withdrawals than every one of the other 17 regions (Mann-Whitney $U = 51$, the largest possible value, $p < 0.001$).

When each new region is added together with its former region, these jumps cancel out (Table 2). In 2024, card cash withdrawals in Zhetysu and Almaty region together grew by only 0.9%, in Abai and East Kazakhstan together by 5.8%, and in Ulytau and Karaganda together by 6.0%. All three are close to the median of the other regions. In 2025 the new regions were already close to normal.

Table 2. Growth in 2024 in the regions created in 2022 and in the regions they were separated from

Region	Card cash withdrawals, %	Cashless payments, %	Household loans, %
Zhetysu	+74.2	+63.5	+49.7
Almaty region	-32.2	+19.0	+26.9
Together	+0.9	+31.5	+40.0
Abai	+29.5	+25.9	+33.4
East Kazakhstan	-6.2	+10.3	+19.0
Together	+5.8	+16.5	+24.3
Ulytau	+29.7	+16.1	+23.4
Karaganda	+2.3	+7.2	+18.8
Together	+6.0	+8.4	+19.4
Median of the other 14 regions	+6.9	+14.4	+22.3

"Together" is the growth of the sum of the two regions. Author's calculations based on National Bank of Kazakhstan data.

This pattern suggests that in 2024 part of the card activity was still being moved between the old and the new regions in the statistics, for example as cards, accounts or cash machines were counted in the new region instead of the old one. The data do not show how exactly this happened, so this is a likely explanation and not a proven fact. It also does not explain everything. Zhetysu and Almaty region together still had faster growth than the other regions in cashless payments (31.5% against a median of 14.4%) and in household loans (40.0% against 22.3%), so the area of the former Almaty region really did grow fast in 2024. Either way, the strong correlation for 2024 depended on how the activity of one area was divided between two regions.

3.4 Taking the boundary change into account

Table 3 shows the correlations with the two corrections for the 2022 boundary change. Without the three new regions, the 2024 correlation between growth in cashless payments and growth in household loans is 0.05 ($p = 0.84$). With the new regions merged back into their former regions, it is 0.26 ($p = 0.32$). Once the boundary change is taken into account, none of the correlations in Table 3 is significant at the 5% level in either year. The strongest one left is between growth in cashless payments and growth in consumer loans in 2024 for the merged regions, $r = 0.42$ ($p = 0.09$), and it drops to 0.26 when the merged Almaty region is removed. The Spearman coefficients lead to the same conclusions. Using the value of loans issued during the year instead of balances, which is possible only for 2025, gives the same result: none of the four correlations is significant (all $p > 0.07$).

Table 3. Pearson correlations across regions under three ways of treating the 2022 boundary change

Measure	Regions	2024, r (p)	2025, r (p)
Cashless growth and household loan growth	All 20	0.66 (0.002)	0.27 (0.242)
	Without new regions (17)	0.05 (0.838)	0.09 (0.718)
	New regions merged (17)	0.26 (0.321)	0.24 (0.361)

Cashless growth and consumer loan growth	All 20	0.70 (0.001)	0.27 (0.256)
	Without new regions (17)	0.31 (0.232)	0.23 (0.379)
	New regions merged (17)	0.42 (0.091)	0.21 (0.417)
Change in cashless share and consumer loan growth	All 20	-0.16 (0.504)	-0.20 (0.407)
	Without new regions (17)	0.17 (0.522)	-0.21 (0.429)
	New regions merged (17)	0.29 (0.251)	-0.14 (0.592)

"New regions merged" means that Zhetysu is added to Almaty region, Abai to East Kazakhstan and Ulytau to Karaganda. Author's calculations.

3.5 The cashless share and consumer loans

The measure that is closest to the research question is the change in the cashless share, because it shows how much a region actually moved from cash to cards and not only how much its card payments grew. This measure was not related to consumer loan growth in any version of the analysis (Table 3, Figure 4). In 2024 the correlation was -0.16 ($p = 0.50$), and in 2025 it was -0.20 ($p = 0.41$). The level of the cashless share at the start of the year was not related to consumer loan growth either ($r = -0.09$, $p = 0.71$ for 2024; $r = 0.36$, $p = 0.12$ for 2025).

Two regions show this well. In 2024, Turkestan and Akmola regions moved away from cash more than almost any other region: their cashless shares rose by 5.6 and 5.8 percentage points, while the median change was 0.7. (The only larger increase, in Almaty region, is affected by the boundary change described in Section 3.3.) Consumer loans in these two regions grew by 33.4% and 35.6%, very close to the median of 35.4%. The fastest move to cashless payments did not come with any unusual growth in borrowing.

4. Discussion

The main answer to the research question is negative. Across the regions of Kazakhstan, a faster move away from cash did not go together with faster growth in household or consumer loans in 2024 or 2025. The only strong correlation in the data, between growth in cashless payments and growth in loans in 2024, depends on one region and on how activity was divided between old and new regions after 2022.

Looking back, it is not surprising that a correlation like $r = 0.66$ appeared. Both variables are measured in money, and both grow when a region's economy, prices and population grow, so a region that grows fast in general will be high on both axes. With only 20 regions, one such region is enough to create a strong correlation. This is the situation Anscombe described when he showed that the same correlation coefficient can come from very different data [19]. For me the practical lesson was that a correlation coefficient is not a result on its own. It has to be checked with a plot and with a leave-one-out test before it is interpreted.

The negative result does not mean that the card effect found in experiments [6, 7] is absent in Kazakhstan. Regional data cannot answer that question. A person may spend more when paying by card, while regional totals are driven mostly by other things, such as incomes, prices, the spread of banking apps and national lending conditions. Robinson made the general point that correlations between group figures do not show how individuals behave [14]. In Kazakhstan there is also little variation left to work with: in most regions more than 80% of card turnover is

already cashless, and in 2024 the median region moved by less than one percentage point. Household loans, on the other hand, grew by at least 15.8% in every region in 2024, which suggests that the main reasons for the growth in lending were common to the whole country. Studies that do find a link between digital finance and consumption, such as the study by Li, Wu and Xiao for China [12], use data on thousands of households, which gives them far more statistical power than 20 regions and two years.

A second finding of this paper is about the data. In 2024, two years after the new regions were formed, the regional card statistics still showed large shifts between the new regions and their former regions. Any study that compares growth by region in 2023 and 2024 can be affected, especially a study that pays attention to single regions. Merging each new region with its former region, as in Section 3.4, is a simple way to avoid this problem.

5. Conclusion, limitations and recommendations

This study looked at whether the move from cash to cashless payments in the regions of Kazakhstan was accompanied by faster growth in bank loans to households. A strong correlation appeared for 2024 ($r = 0.66$), but it depended on one region, Zhetysu, and disappeared once the 2022 boundary change was taken into account. The change in the cashless share of card turnover, which measures the move away from cash directly, was not related to consumer loan growth in either year. At the level of regions, there is no evidence that the shift to cashless payments was driving household borrowing in 2024 and 2025.

Limitations

Small sample. Twenty regions and two years give little statistical power, so a real correlation of moderate size could be missed.

Cash is measured only in part. The data include only cash that was taken out with cards. Cash received as wages or from other people and spent without a card is not counted.

Balances and flows. Loans are balances at a date, while payments are totals over a year. Loans issued during the year are a closer match to payments, but in the files used they are available for 2024 and 2025 only, so this check was possible for 2025 alone.

Regional attribution. The publications do not explain how banks assign card operations and loans to regions. Section 3.3 shows that this assignment changed for the new regions.

Nominal values. All values are in current tenge and are not adjusted for inflation. Inflation differs between regions to some extent, which adds noise to the growth rates.

Revisions. The National Bank revises December data after the end of the year [20]. The values used here are the ones available in September 2026.

Correlation only. Even a significant correlation would not show that one variable causes the other.

Recommendations

For researchers who use Kazakhstan's regional statistics for 2023 and 2024: check every result for dependence on single regions, and treat Abai, Zhetysu and Ulytau with care, for example by merging them with their former regions. A short methodological note from the National Bank on how card operations were reassigned to the new regions would help users of these tables.

For future work on this question, data on individual households or bank customers would be needed to see whether people who switch to cards also borrow more. Longer regional series, after the statistics for the new regions settle, would also make a regional test more reliable.

Data availability

The National Bank releases used in this study are publicly available at the links given in the references. The regional loan files and the spreadsheet with all calculations are available from the author on request. I could not find a stable public link to the regional loan files on the National Bank website.

AI-use disclosure

The author collected the data from the National Bank publications, built the spreadsheets and made the first correlation and leave-one-out calculations in Microsoft Excel. Generative AI tools were used for language editing, for structural revision of the text, and to help re-check the calculations and produce the figures in Python. All numbers in the paper were checked against the original National Bank publications.

Conflicts of interest

The author declares no conflicts of interest.

References:

[1] National Bank of the Republic of Kazakhstan. (2024). Payment cards and electronic banking services. Version No. 244 of 29.01.2024 [in Russian]. https://nationalbank.kz/ru/news/elektronnye-bankovskie-uslugi/16315

[2] National Bank of the Republic of Kazakhstan. (2026). Payment cards and electronic banking services. Version No. 268 of 30.01.2026 [in Russian]. https://nationalbank.kz/ru/news/elektronnye-bankovskie-uslugi/18537

[3] Klapper, L., Singer, D., Starita, L., & Norris, A. (2025). The Global Findex Database 2025: Connectivity and financial inclusion in the digital economy. World Bank. https://doi.org/10.1596/978-1-4648-2204-9

[4] National Bank of the Republic of Kazakhstan. (2025, 2026). Loans of the banking sector to the population by region, 2024 and 2025 [data files, in Russian: "Kredity bankovskogo sektora naseleniyu v regionalnom razreze"].

[5] Feinberg, R. A. (1986). Credit cards as spending facilitating stimuli: A conditioning interpretation. Journal of Consumer Research, 13(3), 348-356. https://doi.org/10.1086/209074

[6] Prelec, D., & Simester, D. (2001). Always leave home without it: A further investigation of the credit-card effect on willingness to pay. Marketing Letters, 12(1), 5-12. https://doi.org/10.1023/A:1008196717017

[7] Soman, D. (2001). Effects of payment mechanism on spending behavior: The role of rehearsal and immediacy of payments. Journal of Consumer Research, 27(4), 460-474. https://doi.org/10.1086/319621

[8] Raghubir, P., & Srivastava, J. (2008). Monopoly money: The effect of payment coupling and form on spending behavior. Journal of Experimental Psychology: Applied, 14(3), 213-225. https://doi.org/10.1037/1076-898X.14.3.213

[9] Prelec, D., & Loewenstein, G. (1998). The red and the black: Mental accounting of savings and debt. Marketing Science, 17(1), 4-28. https://doi.org/10.1287/mksc.17.1.4

[10] Thomas, M., Desai, K. K., & Seenivasan, S. (2011). How credit card payments increase unhealthy food purchases: Visceral regulation of vices. Journal of Consumer Research, 38(1), 126-139. https://doi.org/10.1086/657331

[11] Humphrey, D. B. (2004). Replacement of cash by cards in US consumer payments. Journal of Economics and Business, 56(3), 211-225. https://doi.org/10.1016/j.jeconbus.2003.09.001

[12] Li, J., Wu, Y., & Xiao, J. J. (2020). The impact of digital finance on household consumption: Evidence from China. Economic Modelling, 86, 317-326. https://doi.org/10.1016/j.econmod.2019.09.027

[13] Brito, D. L., & Hartley, P. R. (1995). Consumer rationality and credit cards. Journal of Political Economy, 103(2), 400-433. https://doi.org/10.1086/261988

[14] Robinson, W. S. (2009). Ecological correlations and the behavior of individuals. International Journal of Epidemiology, 38(2), 337-341. https://doi.org/10.1093/ije/dyn357 (Original work published 1950 in American Sociological Review, 15(3))

[15] National Bank of the Republic of Kazakhstan. (2025). Payment cards and electronic banking services. Version No. 256 of 30.01.2025 [in Russian]. https://nationalbank.kz/ru/news/elektronnye-bankovskie-uslugi/17371

[16] National Bank of the Republic of Kazakhstan. (n.d.). Loans of the banking sector to the economy [in Russian]. Retrieved September 26, 2026, from https://nationalbank.kz/ru/news/kredity-bankovskogo-sektora-ekonomike

[17] Decree of the President of the Republic of Kazakhstan No. 887 of 3 May 2022 "On some issues of the administrative-territorial structure of the Republic of Kazakhstan" [in Russian]. https://adilet.zan.kz/rus/docs/U2200000887

[18] Cook, R. D. (1977). Detection of influential observation in linear regression. Technometrics, 19(1), 15-18. https://doi.org/10.1080/00401706.1977.10489493

[19] Anscombe, F. J. (1973). Graphs in statistical analysis. The American Statistician, 27(1), 17-21. https://doi.org/10.1080/00031305.1973.10478966

[20] National Bank of the Republic of Kazakhstan. (n.d.). Monetary and banking statistics [in Russian]. Retrieved September 26, 2026, from https://nationalbank.kz/ru/page/denezhno-kreditnaya-statistika

Figures

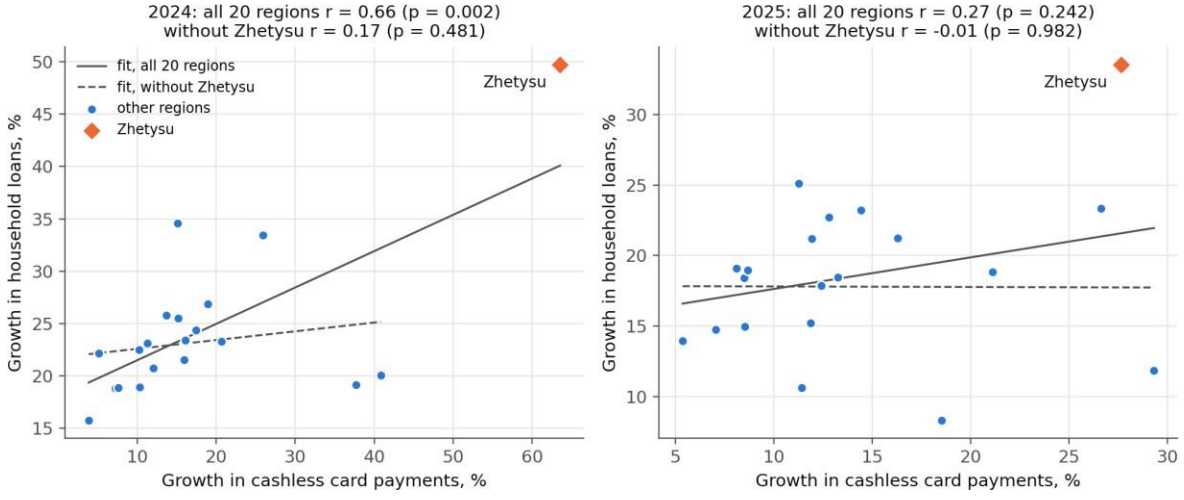


Figure 1. Growth in cashless card payments and growth in household loans across 20 regions, 2024 and 2025. Lines show linear fits with and without Zhetysu.

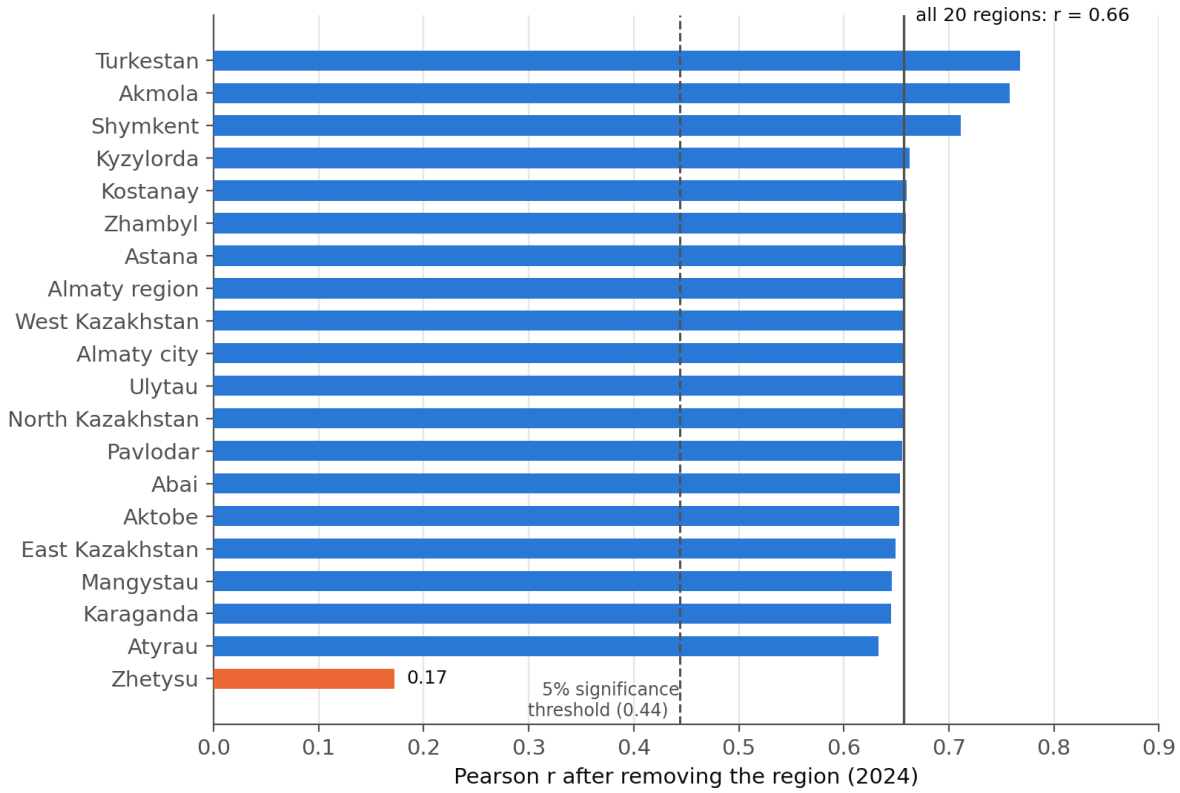


Figure 2. Leave-one-out check for 2024: the correlation between growth in cashless payments and growth in household loans after removing each region in turn.

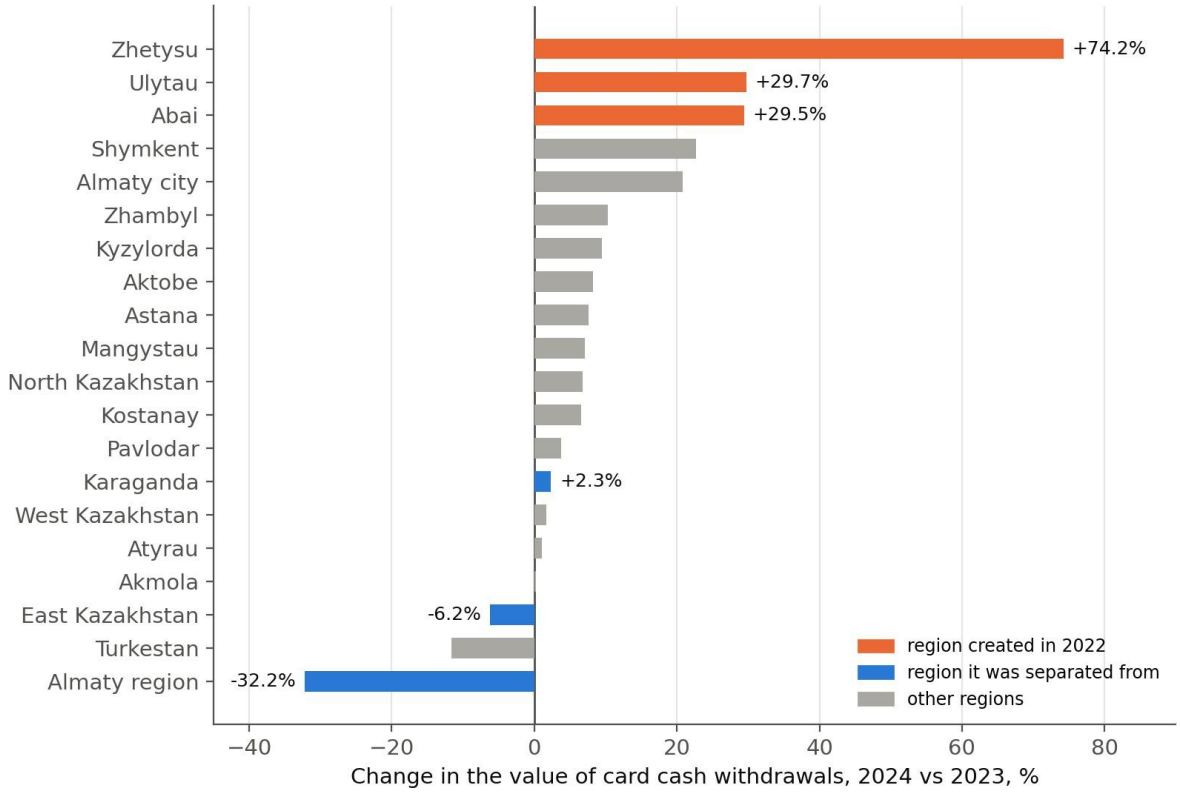


Figure 3. Change in the value of card cash withdrawals in 2024 compared with 2023, by region.

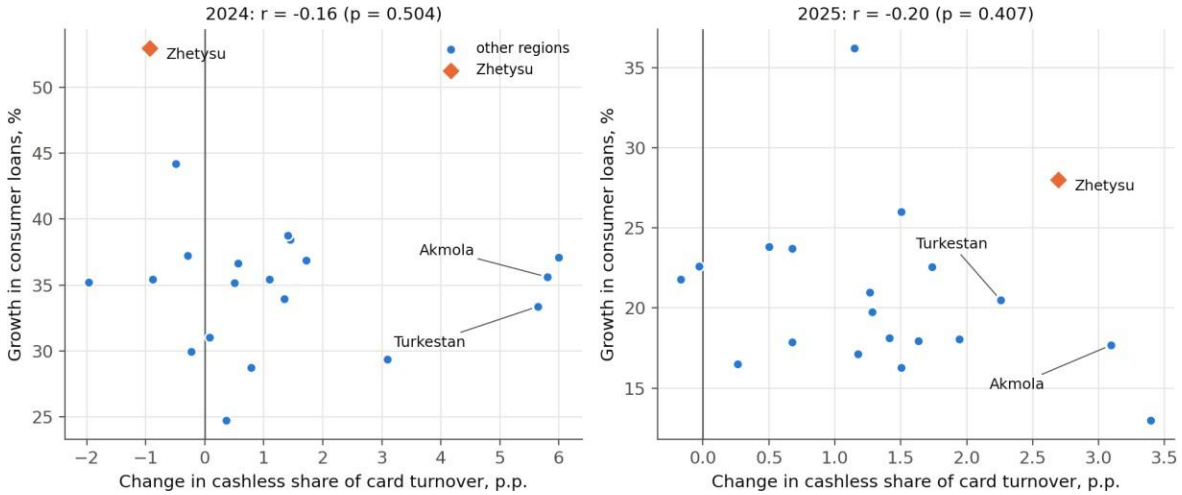


Figure 4. Change in the cashless share of card turnover and growth in consumer loans across 20 regions, 2024 and 2025.

UDC 338.5

Kazhgerey Asan
Grade 11 student
Supervisor: Askhat Serikbekuly
Republican Physics and Mathematics School
(Astana, Kazakhstan)

HOW MANY FOOD BASKETS DOES AN AVERAGE WAGE BUY? WAGES AND FOOD PRICES IN THE 20 REGIONS OF KAZAKHSTAN, 2024-2026

Abstract: Average wages in Kazakhstan are growing quickly, but so are food prices, and both differ a lot between regions. In this study I asked whether an average monthly wage in each region could buy more or less food in 2026 than two years earlier. I collected official retail prices for 12 basic foods in the 20 regional centres from 32 monthly bulletins of the Bureau of National Statistics (January 2024 to August 2026) and the average monthly nominal wage for the 20 regions (first quarter of 2024 to second quarter of 2026). Milk, pasta and rice were left out of the main analysis because the Bureau changed these items in January 2026. For the other nine foods I built a simple basket with one unit of each and counted how many such baskets one average wage could buy. Between the second quarter of 2024 and the second quarter of 2026, the basket became 27.7% more expensive on average, while wages grew by 19.2%. As a result, the number of baskets per wage fell in 17 of the 20 regions, with the median region losing 8.5%. The largest falls were in Ulytau (-13.1%), Atyrau (-12.6%) and West Kazakhstan (-12.5%), and the largest rise was in Astana (+6.3%). Beef alone explained 57% of the increase in the basket's cost, and without beef the number of baskets fell in only 11 regions. The results show that wage growth did not keep up with the price of basic food in most regions, but also that this conclusion depends strongly on one product.

Keywords: food prices; wages; purchasing power; food basket; regional statistics; Kazakhstan

Introduction. When people in Kazakhstan talk about prices, they usually mean food. Every month the Bureau of National Statistics publishes the prices of basic foods in every regional centre, and every quarter it publishes the average wage in every region [1, 2]. Both numbers have been rising. The national average monthly wage went from 403,251 tenge in the second quarter of 2024 to 486,388 tenge in the second quarter of 2026 [2]. But a higher wage does not mean much by itself. What matters for a family is what the wage can buy, and especially whether it can buy basic food.

Researchers usually measure this by comparing the cost of food with income. International organisations calculate how much a healthy diet costs in each country and what share of people cannot afford it [3, 4]. Bai and co-authors did this for 177 countries using retail prices of individual foods [5]. Reeves, Loopstra and Stuckler compared food prices with wages in 21 European countries and found that when food prices rose faster than wages, more people reported that they could not afford meat, chicken or fish every second day [6]. So the gap between food prices and wages is not only an abstract number, it is connected to what people actually eat.

Kazakhstan is a good place to look at this inside one country, because the regions are very different. In 2024 the average wage in Atyrau region was about twice as high as in Zhetysu region, while basic food cost about the same in both [1, 2]. If food prices and wages grow at

different speeds in different regions, some regions can become poorer in real terms even when their wages grow.

In this study I asked three questions:

• How many baskets of basic foods could an average monthly wage buy in each region of Kazakhstan in the second quarter of 2024 and in the second quarter of 2026?

• In which regions did this number go down, and was the cause slow wage growth or fast price growth?

• Which foods pushed the cost of the basket up the most?

Data and methods. Food prices. I took the prices from the monthly bulletin "Retail prices for selected goods and services in the Republic of Kazakhstan" (form T-15-03-M) of the Bureau of National Statistics [1]. Each bulletin gives the average price of more than 50 foods in the capital, the two other cities of republican significance and the 17 regional centres. I downloaded all 32 bulletins from January 2024 to August 2026 and copied the prices of 12 foods that people buy almost every week: wheat bread made from first-grade flour, top-grade wheat flour, rice, pasta, sunflower oil, beef on the bone, chicken, eggs, sugar, potatoes, onions and milk. This gave 7,680 price records: 20 places, 12 foods and 32 months. For each region I used the price in its regional centre, for example Aktau for Mangystau region and Zhezkazgan for Ulytau region.

Wages. I took the average monthly nominal wage by region and by quarter from the Bureau's indicator 252101 [2]. This gave 200 values: 20 regions and 10 quarters, from the first quarter of 2024 to the second quarter of 2026.

Three foods left out. In the January 2026 bulletin the Bureau changed some of the items. Milk is now described as "milk 3.2-4.5% fat" instead of "pasteurised milk 3.2-4%", so it is a different product. On the same date the national average price of pasta fell from 633 to 495 tenge (-21.8%) and the price of rice fell from 625 to 551 tenge (-11.8%), and in the following months both prices changed only a little [7, 8]. Real prices do not usually move like this in one month, so the change most likely comes from what was being measured, not from the market. The January 2026 bulletin was also published later than usual, with a note that the data had been reformatted because of a technical error [8]. Because I could not compare these three foods before and after January 2026, I left them out of the main basket. In addition, for 2024 and 2025 the Bureau did not publish a milk price for five regional centres (Atyrau, Kyzylorda, Aktau, Turkestan and Shymkent). I kept these cells empty and did not fill them with averages.

The basket. The main basket contains one unit of each of the nine remaining foods: 1 kg of bread, 1 kg of flour, 1 litre of sunflower oil, 1 kg of beef, 1 kg of chicken, 10 eggs, 1 kg of sugar, 1 kg of potatoes and 1 kg of onions. For each quarter I took the average of the three monthly prices of each food and added them up. Then I divided the average wage of the region by the cost of the basket. The result is the number of baskets that one average monthly wage could buy. I call it "baskets per wage".

Comparison. I compared the second quarter of 2024 with the second quarter of 2026. I used the same quarter in both years because wages and some prices change with the season. For example, the average number of baskets per wage across the regions was 59.4 in the second quarter of 2024 but 62.5 in the fourth quarter of 2024, when wages are higher at the end of the year. Comparing the same quarter removes this effect.

Statistics. To check whether 17 falls out of 20 could happen by chance, I used a sign test. To see whether the change in baskets per wage was linked more to wages or to prices, I used

Spearman's rank correlation [9]. I also repeated the main comparison without beef and with the first quarter instead of the second, to see if the result depends on these choices.

Results

Baskets per wage in 2024 and 2026. In the second quarter of 2024, the basket of nine foods cost 6,534 tenge on average across the 20 regional centres. In the second quarter of 2026 it cost 8,347 tenge, which is 27.7% more. Over the same two years the average regional wage grew by 19.2%, from 389,894 to 464,790 tenge. The average number of baskets per wage fell from 59.4 to 55.6.

Table 1 shows every region. Baskets per wage fell in 17 of the 20 regions. If rises and falls were equally likely, a split as uneven as 17 to 3 would happen with a probability of only about 0.3% (two-sided sign test, p = 0.003). The median region lost 8.5% of its baskets. The largest falls were in Ulytau (-13.1%), Atyrau (-12.6%) and West Kazakhstan (-12.5%). Baskets per wage went up only in Astana (+6.3%), Pavlodar (+1.1%) and North Kazakhstan (+0.4%), and in Mangystau (-0.5%) and Akmola (-0.7%) they hardly changed.

Table 1. Wages, basket cost and baskets per wage by region, second quarter of 2024 and 2026

Region	Wage growth, %	Basket growth, %	Baskets per wage, Q2 2024	Baskets per wage, Q2 2026	Change, %
Ulytau	+12.6	+29.5	76.1	66.1	-13.1
Atyrau	+6.2	+21.5	87.6	76.6	-12.6
West Kazakhstan	+17.5	+34.2	58.2	51.0	-12.5
Zhambyl	+16.3	+30.4	50.8	45.3	-10.8
Aktobe	+18.9	+33.1	64.9	58.0	-10.7
Kyzylorda	+16.0	+29.2	58.8	52.8	-10.2
Almaty region	+18.2	+31.6	49.5	44.4	-10.2
Abai	+24.1	+36.2	56.1	51.1	-8.9
Zhetysu	+25.8	+37.7	46.0	42.0	-8.7
East Kazakhstan	+21.2	+32.6	56.5	51.6	-8.6
Shymkent city	+20.2	+31.3	48.2	44.1	-8.4
Turkestan	+14.3	+22.8	51.0	47.5	-6.9
Almaty city	+27.0	+34.9	70.2	66.1	-5.8
Kostanay	+22.2	+25.6	52.8	51.4	-2.7
Karaganda	+24.0	+27.0	59.4	57.9	-2.4
Akmola	+23.0	+23.9	50.5	50.2	-0.7
Mangystau	+16.2	+16.8	78.6	78.2	-0.5
North Kazakhstan	+22.1	+21.7	45.3	45.5	+0.4
Pavlodar	+22.8	+21.5	56.5	57.1	+1.1
Astana city	+25.0	+17.5	70.3	74.8	+6.3

Wages: average monthly nominal wage in the region. Basket: one unit of each of nine foods at prices of the regional centre. Regions are sorted from the largest fall to the largest rise. Source: author's calculations based on data of the Bureau of National Statistics.

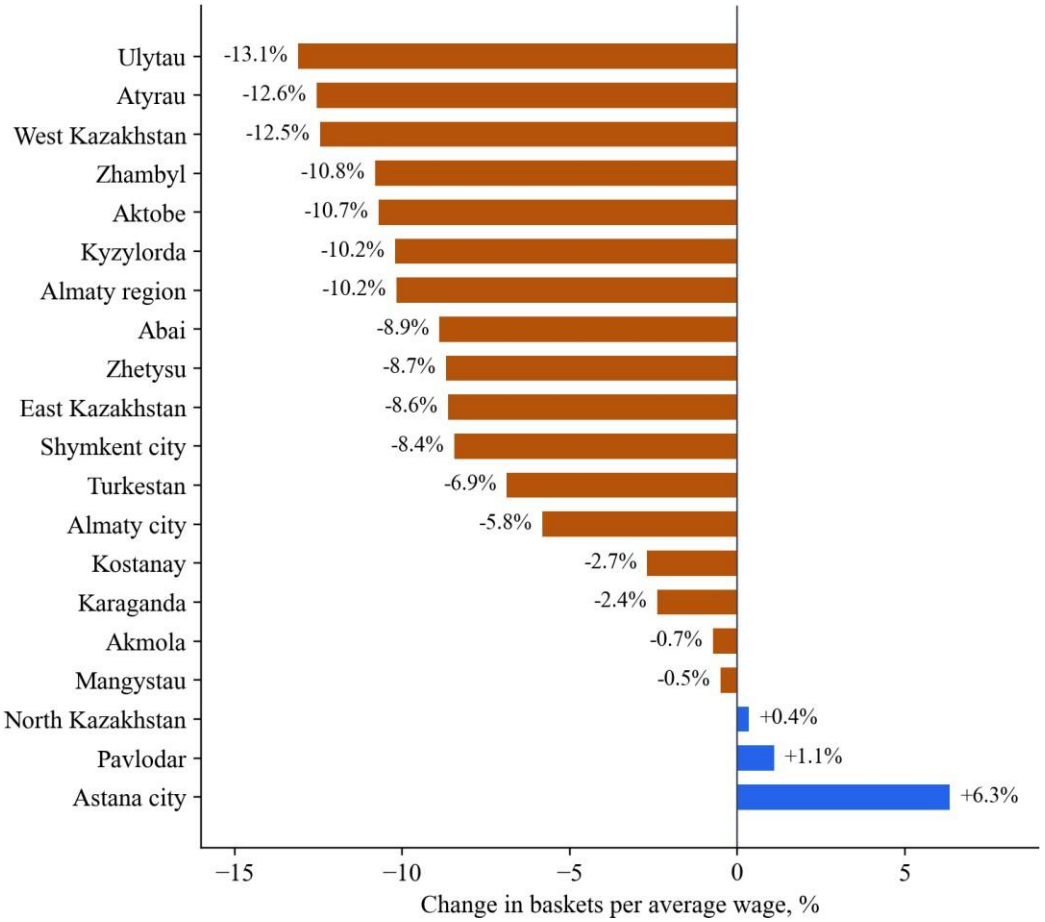


Figure 1. Change in the number of food baskets one average wage can buy, second quarter of 2024 to second quarter of 2026

Wages or prices? The basket became more expensive faster than wages grew in the same 17 regions. The median region's wage grew by 20.7%, and the median basket by 29.4%. Both sides mattered. Across the regions, the change in baskets per wage was linked to wage growth (Spearman's $\rho = 0.57$, $p = 0.008$) and, in the opposite direction, to basket growth ($\rho = -0.53$, $p = 0.015$), and the two links were about equally strong.

Figure 2 shows this. Each point is a region, and the diagonal line marks the regions where wages and the basket grew at the same speed. Most regions are below the line. Astana is the clearest exception: its wages grew by 25.0%, one of the highest rises, while its basket became only 17.5% more expensive. In Ulytau the situation was the opposite: wages grew by 12.6% and the basket by 29.5%. Atyrau and Mangystau had the highest wages in 2024 and both still had wages much higher than most regions in 2026, but Atyrau's wages grew by only 6.2%, the slowest in the country.

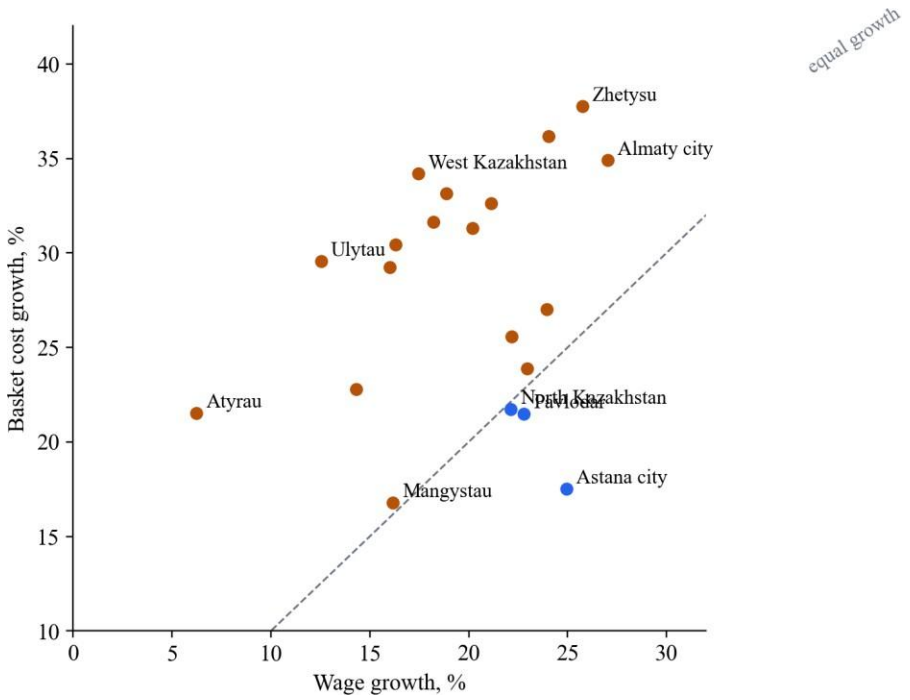


Figure 2. Wage growth and basket growth by region, second quarter of 2024 to second quarter of 2026. Below the line the basket grew faster than wages

Which foods pushed the basket up. Table 2 shows how much each food contributed to the increase in the basket's cost. Beef was the most important by far. It made up 44% of the basket in 2026 and its price rose by 38.8%, so it explained 56.7% of the whole increase. Sunflower oil (+41.0%) explained another 14.4%, and chicken (+16.1%) explained 12.9%. Potatoes and onions became much more expensive in percentage terms (+54.3% and +43.1%), but they are cheap, so they added little to the basket. Sugar changed the least (+4.7%).

Table 2. Average prices of the nine foods across the 20 regional centres and their contribution to the increase in the basket's cost

Food	Price Q2 2024, tenge	Price Q2 2026, tenge	Change, %	Share of the basket increase, %
Beef on the bone, kg	2,648	3,675	+38.8	56.7
Sunflower oil, litre	638	899	+41.0	14.4
Chicken, kg	1,454	1,689	+16.1	12.9
Potatoes, kg	163	252	+54.3	4.9
Eggs (10)	512	581	+13.4	3.8
Onions, kg	113	162	+43.1	2.7
Wheat flour (top grade), kg	384	421	+9.4	2.0
Wheat bread (first-grade flour), kg	203	229	+13.2	1.5
Sugar, kg	419	439	+4.7	1.1

Prices are averages of the three months of the quarter and of the 20 regional centres. Source: author's calculations based on data of the Bureau of National Statistics.

Checking the result. Because beef is such a large part of the basket, I repeated the comparison without it. With only eight foods, baskets per wage still fell in 11 of the 20 regions,

but this is much closer to half. I also compared the first quarter of 2024 with the first quarter of 2026 instead of the second quarters. In that comparison baskets per wage fell in 13 of the 20 regions. So the direction of the result stays the same, but its size depends on the choice of foods and on the quarter. The finding that wages fell behind food prices is strongest for a basket that includes meat.

Discussion. The main result is that between 2024 and 2026 an average wage in most regions of Kazakhstan could buy less basic food than before, even though wages grew by about a fifth. This is similar to what Reeves and co-authors described for Europe, where food prices rose faster than wages [6]. In their study this gap was connected with more people being unable to afford meat or fish regularly. My data cannot show what families actually bought, but the fact that meat was the main reason for the rise in my basket points in the same direction.

The regional differences are also important. The average result for the country hides very different situations. In Astana the wage became worth more food, while in Ulytau, Atyrau and West Kazakhstan it lost more than a tenth. Some regions with high wages, like Atyrau, lost a lot because their wages grew slowly, and other regions lost because their food prices grew quickly. So a single national number for prices or wages is not enough to understand how people in different regions live.

Working on this study I also learned how much the result depends on the data. When the Bureau changed the description of milk and the prices of pasta and rice jumped at the same time, a careless comparison would have shown changes that never happened in the shops. I first filled the missing milk prices with averages from other regions, but this created a 61% "increase" in the price of milk in Atyrau in one month, which was not real. That is why I kept missing values empty and left these three foods out.

Limitations. This study has several limitations. First, the basket has one unit of each food, so it does not show how much of each food a real family buys. Beef has a large weight in my basket simply because it is the most expensive item per kilogram. Second, the prices are measured in the regional centre, while the wages are averages for the whole region, including smaller towns and villages where prices can be different. Third, the average wage is pulled up by a small number of very high salaries, especially in oil regions like Atyrau and Mangystau, so it is higher than what a typical worker earns. Fourth, I used nominal wages and did not include pensions, social payments or other income. Finally, two years is a short period, and the result is weaker when the first quarter is used instead of the second.

Conclusion. Between the second quarter of 2024 and the second quarter of 2026, a basket of nine basic foods became 27.7% more expensive across the regional centres of Kazakhstan, while wages grew by 19.2%. The number of baskets an average wage could buy fell in 17 of the 20 regions, most of all in Ulytau, Atyrau and West Kazakhstan, and rose only in Astana, Pavlodar and North Kazakhstan. More than half of the increase in the basket's cost came from beef. Future studies could use a basket that matches what families actually buy and add data from the Bureau's household budget surveys to see whether the same regions lost purchasing power in real spending.

Author contribution and use of AI. I collected the data myself: I downloaded all 32 monthly price bulletins and the regional wage table from the website of the Bureau of National Statistics, built the price and wage tables (7,680 and 200 records), and calculated the basket and the number of baskets per wage. My mentor checked all the prices and wages against the original files of the Bureau, ran the statistical tests and helped me write the text in English. No data,

number or reference in this paper was produced by an AI tool; all numbers come from the official data and the calculations described in Section 2.

Data availability. All data come from the public website of the Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan. The price and wage tables used in this study are available from the author on request.

References:

1. Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan. Retail prices for selected goods and services in the Republic of Kazakhstan (form T-15-03-M), monthly bulletins for January 2024 - August 2026 [in Russian]. URL: https://stat.gov.kz/ru/industries/economy/prices/spreadsheets/?name=19087 (accessed: 01.10.2026).

2. Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan. Average monthly nominal wage by regions of the Republic of Kazakhstan (indicator 252101), dynamic table, updated 10.09.2026 [in Russian]. URL: https://stat.gov.kz/ru/industries/labor-and-income/stat-wags/dynamic-tables/ (accessed: 01.10.2026).

3. FAO, IFAD, UNICEF, WFP, WHO. The State of Food Security and Nutrition in the World 2024. Rome: FAO, 2024. DOI: 10.4060/cd1254en.

4. Herforth A., Bai Y., Venkat A., Mahrt K., Ebel A., Masters W.A. Cost and affordability of healthy diets across and within countries. FAO Agricultural Development Economics Technical Study No. 9. Rome: FAO, 2020. DOI: 10.4060/cb2431en.

5. Bai Y., Alemu R., Block S.A., Headey D., Masters W.A. Cost and affordability of nutritious diets at retail prices: Evidence from 177 countries // Food Policy. 2021. Vol. 99. Art. 101983. DOI: 10.1016/j.foodpol.2020.101983.

6. Reeves A., Loopstra R., Stuckler D. The growing disconnect between food prices and wages in Europe: cross-national analysis of food deprivation and welfare regimes in twenty-one EU countries, 2004-2012 // Public Health Nutrition. 2017. Vol. 20, No. 8. P. 1414-1422. DOI: 10.1017/S1368980017000167.

7. Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan. Retail prices for selected goods and services in the Republic of Kazakhstan, December 2025. Published 09.01.2026 [in Russian]. URL: https://stat.gov.kz/ru/industries/economy/prices/spreadsheets/?name=19087 (accessed: 01.10.2026).

8. Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan. Retail prices for selected goods and services in the Republic of Kazakhstan, January 2026. Published 05.03.2026 [in Russian]. URL: https://stat.gov.kz/ru/industries/economy/prices/spreadsheets/?name=19087 (accessed: 01.10.2026).

9. Spearman C. The proof and measurement of association between two things // The American Journal of Psychology. 1904. Vol. 15, No. 1. P. 72-101. DOI: 10.2307/1412159.

UDC 336.7

Tagimova Alua
Grade 11 student
Scientific supervisor: Altynbek Aruzhan Yessenqyzy
Aktobe Boarding School for Girls
(Aktobe, Kazakhstan)

WHO WAS LEFT BEHIND? EDUCATION AND FINANCIAL INCLUSION IN KAZAKHSTAN AND THREE CENTRAL ASIAN NEIGHBOURS, 2021-2024

Abstract: By 2024, 87% of adults in Kazakhstan had a financial account, more than in any of its Central Asian neighbours. I used the World Bank's Global Findex data for 2021 and 2024 to check whether this growth reached adults with only primary education or less, and compared Kazakhstan with Kyrgyzstan, Tajikistan and Uzbekistan. In Kazakhstan, account ownership in this group did not grow (61.9% in 2021 and 58.1% in 2024), while among adults with secondary education or more it rose from 84.3% to 91.2%. The same was true for digital payments and for saving any money. In the three neighbouring countries, account ownership among the less educated grew by 11 to 22 percentage points. The less educated group in Kazakhstan did change in one way: those who saved moved their savings into banks. The share of savers in this group who saved at a bank or similar institution rose from about 15% to about 47%, while in the more educated group it stayed at about 52%. The less educated group is small, around 13% of the sample, so the results have wide margins of error, and the widening of the gap between the two groups is not statistically certain. Still, the evidence suggests that Kazakhstan's high national average hides a group that the recent growth in financial services did not reach.

Keywords: financial inclusion; education; account ownership; saving; Global Findex; Kazakhstan; Central Asia

Introduction

Kazakhstan looks like a success story in financial inclusion. According to the World Bank's Global Findex database, the share of adults with an account at a bank, a similar institution or a mobile money provider grew from 58.7% in 2017 to 81.1% in 2021 and 87.0% in 2024 [1]. This is the highest level in Central Asia: in 2024 the figure was 72.3% in Kyrgyzstan, 59.7% in Uzbekistan and 54.5% in Tajikistan [1].

What I wanted to know is who is behind the average. A national figure of 87% can hide groups that are still far from the rest. Research on financial inclusion has shown for a long time that account ownership differs not only between countries but also between income groups inside the same country [2]. In a study of 37 African countries based on the same Findex survey, Zins and Weill found that education and income were the strongest individual factors behind financial inclusion [3]. Allen and co-authors showed that account ownership is higher where accounts are cheaper and financial institutions are closer, and that policies to lower these barriers may be especially effective for the groups most likely to be excluded [4]. There is also a knowledge side. Lusardi and Mitchell describe financial knowledge as a kind of human capital that is unevenly spread across the population [5], and a study by Cole, Sampson and Zia that combined surveys with a field experiment found that a financial education programme increased

demand for bank accounts only among people with limited education or financial literacy, while small subsidies had a much larger effect [6].

For Kazakhstan, the research is thinner. A recent review of financial inclusion in the country notes that there are few comprehensive studies of these processes in Kazakhstan [7]. The newest Findex round, collected in 2024 and published in 2025 [8], makes it possible to look at changes since the previous round in 2021 [9] and to split the results by education.

In this study I compare two education groups that Findex reports for every country: adults with primary education or less, and adults with secondary education or more. I ask three questions:

- Did account ownership among adults with primary education or less grow in Kazakhstan between 2021 and 2024, as it did among more educated adults?
- Did the way people in each group save money change, in particular the share who save at a bank?
- Is Kazakhstan's pattern the same as in Kyrgyzstan, Tajikistan and Uzbekistan?

Materials and Methods

Data. All data come from the Global Findex Database 2025, which is published by the World Bank and includes the 2017, 2021 and 2024 survey rounds [1, 8]. The survey is carried out by Gallup as part of the Gallup World Poll and interviews about 1,000 randomly selected people aged 15 and older in each economy, using nationally representative samples [10]. I did not use individual answers, only the published percentages for each country and group. I collected the numbers from the DataBank website, and they were then checked a second time against the same data downloaded through the World Bank's data interface (source 28). All values matched.

Indicators. I used five indicators, each reported separately for adults with primary education or less and for adults with secondary education or more: having an account; having made or received a digital payment; having saved any money in the past year; having saved at a bank or similar financial institution; and having saved at a financial institution or using a mobile money account. For Kazakhstan I also used the share with an account at a bank or similar institution only. The countries were Kazakhstan, Kyrgyzstan, Tajikistan and Uzbekistan, and the years were 2021 and 2024. For the three neighbours, the saving-at-a-bank indicators by education are missing in most cells, so for them I compare only account ownership and saving any money.

One detail of the data turned out to be important. For Kazakhstan, the database has no mobile money values for 2021, and in that year the share with "an account" is exactly the same as the share with a bank account (81.1%). In 2024 the account figure includes mobile money accounts. The same applies to saving: in 2021, "saved at a financial institution or using mobile money" equals saving at a bank, while in 2024 it also includes mobile money. Because of this, when I compare saving in 2021 and 2024 I use saving at a bank or similar institution, which is measured the same way in both years, and I show saving through mobile money separately.

Calculations. For each indicator I calculated the gap between the two groups (secondary or more minus primary or less) and the change from 2021 to 2024. For saving, I calculated what share of all savers in each group saved at a bank: saving at a bank divided by saving any money. All differences and shares were calculated from the unrounded values in the database and rounded only at the end, so a difference in the text can differ by 0.1 from the difference of the rounded numbers.

Margins of error. The database gives only the percentages, without margins of error, so I estimated them roughly. The less educated group is small. Comparing the national average with the two group averages shows that it makes up about 14% of the Kazakhstan sample in 2021 and about 12-13% in 2024, which is roughly 120-140 people. For each change I estimated a 95% margin of error as if the group were a simple random sample of that size. The real margins are somewhat wider, because the survey uses weights and is not a simple random sample, so I treat these numbers as a rough guide and call a change clear only when it is well outside the margin.

Results

Account ownership in Kazakhstan

The biggest difference in Kazakhstan was between the two education groups (Table 1). In 2021, 61.9% of people with primary education or less had an account, compared with 84.3% of people with secondary education or more. In 2024 the figures were 58.1% and 91.2%, so the gap grew from 22.4 to 33.2 percentage points. The change from 61.9% to 58.1% should not be treated as a

clear decline. This group is small, and the 3.8-point change is well within the rough margin of error of about ±12 points. So it is more accurate to say that account ownership did not grow for this group, while it clearly increased for the more educated group (+7.0 points, margin about ±3). For the same reason, the widening of the gap by 10.8 points is also within the margin of about ±12 points and cannot be called certain.

Digital payments followed the same pattern. The share of less educated adults who made or received a digital payment stayed almost the same (55.8% and 55.3%), while among more educated adults it rose from 81.6% to 89.3%. It is interesting that this group had made a very large jump earlier: its account ownership was only 29.6% in 2017 and more than doubled by 2021. The growth stopped after 2021.

Table 1. Financial inclusion in Kazakhstan by education group, 2021 and 2024

Indicator, % of adults in the group	Primary or less, 2021	Primary or less, 2024	Secondary or more, 2021	Secondary or more, 2024
Has an account	61.9	58.1	84.3	91.2
Made or received a digital payment	55.8	55.3	81.6	89.3
Saved any money	31.4	32.2	41.6	55.7
Saved at a bank or similar institution	4.7	15.0	21.4	29.0
Saved at a bank or through mobile money	4.7*	16.5	21.4*	40.7

Source: Global Findex Database 2025. *Mobile money was not measured for Kazakhstan in 2021, so in that year this row equals saving at a bank.

Where people save

The savings numbers were more interesting to me. Among people with primary education or less, the share saving any money barely changed, from 31.4% to 32.2%. But saving at a bank or similar institution increased from 4.7% to 15.0%, which is a clear change (margin about ±7 points). In 2021, only about 15% of savers in this group (4.7 out of 31.4) saved at a bank; in 2024 it was about 47%. So the main change was not that more people started saving, but that

more of the people who already saved moved their savings into the formal system. Mobile money added very little for this group: only 1.4% of adults in it saved through mobile money without also saving at a bank.

The more educated group showed a different pattern (Figure 1). The share saving any money increased from 41.6% to 55.7%, and saving at a bank went up from 21.4% to 29.0%. But the share of savers who saved at a bank stayed almost the same, about 52% in both years. Bank saving grew at the same speed as saving in general. What was new in 2024 was mobile money: 11.7% of adults in this group saved through mobile money without saving at a bank. If mobile money is counted, the share of savers who kept their money outside any formal channel fell from 48.4% to 27.0% in this group, and from 85.2% to 48.9% in the less educated group. Because mobile money was not measured in 2021, I think the bank-only comparison is the fairer one, and on that measure only the less educated group changed where it saves.

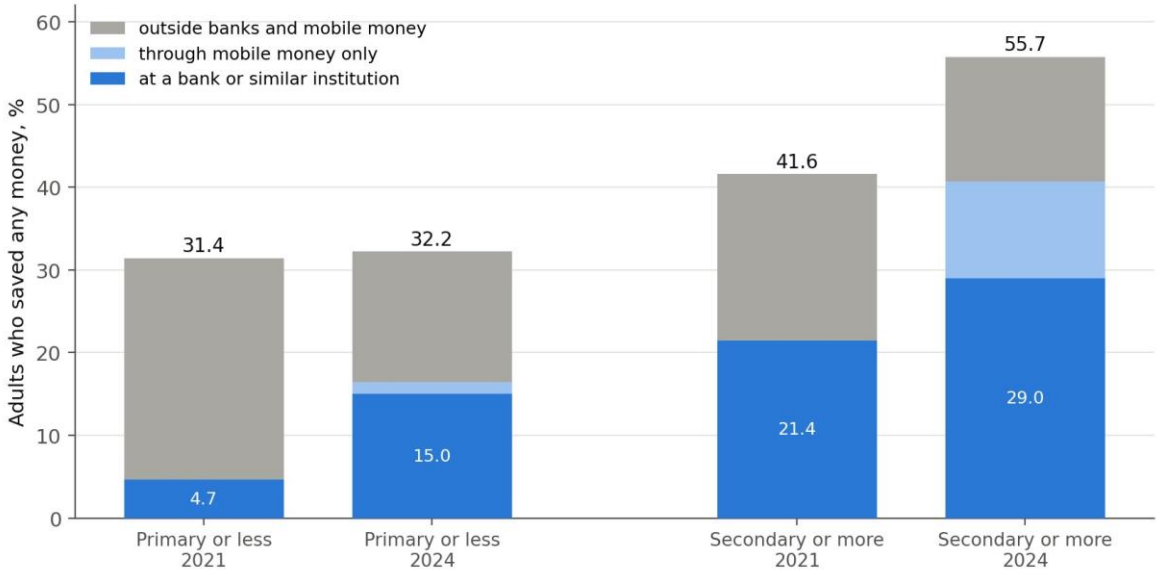


Figure 1. Kazakhstan: adults who saved any money, by where they saved, education group and year. Mobile money was not measured in 2021, so any mobile money saving in that year is included in the grey part.

Comparison with the neighbouring countries

When I compared Kazakhstan with the other three countries, the difference became clearer (Table 2, Figure 2). Among people with primary education or less, account ownership grew by 21.5 points in Kyrgyzstan, 15.4 in Uzbekistan and 11.4 in Tajikistan, and in all three countries the growth was larger than the rough margin of error. Kazakhstan did not show the same growth, although it still had the highest account ownership for this group in 2024. This means that Kazakhstan started from a higher level, but its less educated group did not make the same progress over these three years. Uzbekistan was different in one more way: it was the only country where the education gap in account ownership became smaller, from 28.0 to 24.6 percentage points, although this change is also within the margin of error.

Table 2. Account ownership by education group in four Central Asian countries, % of adults

Country	Primary 2021	Primary 2024	Change	Secondary 2021	Secondary 2024	Change	Gap 2021	Gap 2024
Kazakhstan	61.9	58.1	-3.8 (±11.9)	84.3	91.2	+7.0	22.4	33.2
Kyrgyzstan	28.9	50.5	+21.5 (±10.4)	48.9	75.9	+27.0	19.9	25.4
Tajikistan	32.2	43.6	+11.4 (±8.1)	42.3	58.6	+16.3	10.1	15.0
Uzbekistan	23.9	39.3	+15.4 (±8.9)	51.8	63.9	+12.0	28.0	24.6

Source: Global Findex Database 2025. Primary = primary education or less; secondary = secondary education or more; gap = secondary minus primary. Changes and gaps are calculated from unrounded values. The numbers in brackets are rough 95% margins of error for the change in the less educated group.

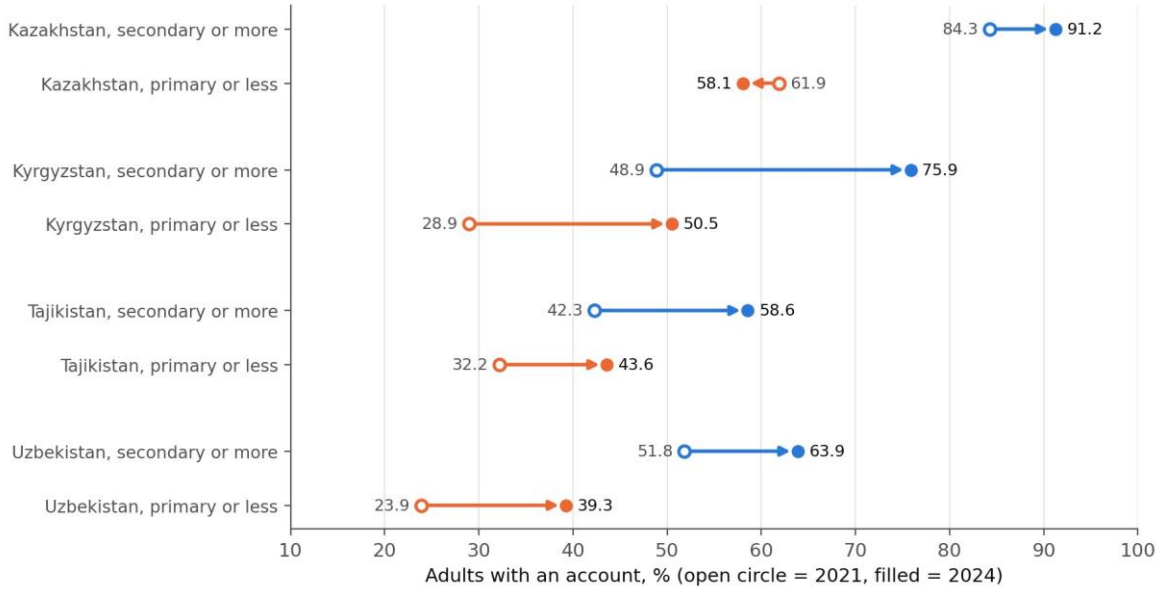


Figure 2. Account ownership by education group, 2021 (open circles) and 2024 (filled circles).

Saving any money was a different story. Among adults with primary education or less it hardly changed anywhere: from 23.7% to 26.1% in Kyrgyzstan, from 16.0% to 21.6% in Tajikistan and from 19.8% to 18.9% in Uzbekistan, and none of these changes is outside the margin of error. So in saving, Kazakhstan's less educated adults were not behind their neighbours. The difference was specific to accounts and payments.

Discussion

The main result of this study is that the recent growth of financial inclusion in Kazakhstan did not reach adults with primary education or less. Between 2021 and 2024 their account ownership, digital payments and saving all stayed at about the same level, while the more educated majority kept growing. In the three neighbouring countries, the less educated group grew clearly over the same years. This fits earlier research that finds education to be one of the strongest individual factors in financial inclusion [3], but the comparison adds something: in Kazakhstan the less educated group already had a higher level than in the neighbouring countries, and it stopped growing, while in the neighbours it was still catching up.

I can think of a few possible explanations, but the data used here cannot test them. One reason could be that Kazakhstan already had a relatively high percentage of people with primary education using banking services. Those who still did not have accounts may have been harder to reach because of limited information, difficulties using banking services or a lack of trust. In Kyrgyzstan, Tajikistan and Uzbekistan the initial percentage was much lower, so there were more people who were only beginning to use bank accounts as banking became more common. This may explain why growth among people with primary education was more noticeable there, while Kazakhstan saw little change. The less educated group in Kazakhstan is also small and may be different from the same group in other countries. It is about 13% of the sample in Kazakhstan, compared with about 27% in Tajikistan and 27-29% in Uzbekistan in 2021, so in Kazakhstan the people who remain in this group could be the ones who are hardest to reach for other reasons, such as age or where they live. Another is that financial services in Kazakhstan now work mostly through apps, which need a certain amount of knowledge and confidence. Research on financial literacy shows that this knowledge is spread unevenly [5], and the experiment by Cole and co-authors suggests that for people with little education both knowledge and costs matter [6]. Studies of the barriers people report, such as cost, distance and documents [2, 4], could help check these ideas for Kazakhstan.

The saving results give a more positive picture. Less educated adults in Kazakhstan did not start saving more, but those who saved were much more likely to use a bank in 2024 than in 2021. For people who do save, this is a real change, because money in a bank is usually safer than cash kept at home. At the same time, their share of bank accounts alone fell from 61.9% to 46.1% while their total account ownership, including mobile money, only moved from 61.9% to 58.1%. I do not want to read too much into this: the 2024 data include mobile money accounts for Kazakhstan while the 2021 data do not, and some people may have described the same account differently in the two rounds. It shows, though, that results for Kazakhstan need to be compared carefully across these two survey rounds.

This study has several limitations. The less educated group is only about 12-14% of the sample, roughly 120-140 people per round, so small changes cannot be detected, and my margins of error are rough and probably too narrow. I only used published percentages, so I could not control for age, income, gender or region, which are connected with education. Mobile money was not measured for Kazakhstan in 2021, which limits the comparison of account ownership and formal saving. Finally, this is a comparison of two years and shows what changed, not why.

Conclusion

Kazakhstan has the highest account ownership in Central Asia, but between 2021 and 2024 this growth did not reach adults with primary education or less. Their account ownership, digital payments and saving stayed flat, while they grew among more educated adults in Kazakhstan and among less educated adults in Kyrgyzstan, Tajikistan and Uzbekistan. The one clear change in this group was that people who save moved much of their savings into banks. For policy, this suggests looking beyond the national average: a small group can be left behind even when the country as a whole looks close to full inclusion. A study with individual Findex data, which could control for age, income and region, would be the natural next step.

Data availability

All data are publicly available from the World Bank Global Findex Database 2025 [1]. The table of all values used in this paper and the script that calculates every number are available from the author on request.

AI and tools disclosure

I collected the data from the Global Findex database and made the first calculations myself. Generative AI tools were used for language editing and structural revision of the text, and to help re-check the calculations and prepare the figures in Python. No data, results or citations were generated by AI: all numbers come from the World Bank database and were checked against it, and all cited sources were checked against the originals. I have reviewed the final manuscript and take full responsibility for it.

Acknowledgements

I thank my supervisor, Altynbek Aruzhan Yessenqyzy, for supporting this work.

References:

1. World Bank. Global Findex Database 2025 [Electronic resource]. URL: https://databank.worldbank.org/source/global-financial-inclusion (accessed: 28.09.2026).

2. Demirgüç-Kunt A., Klapper L. Measuring financial inclusion: Explaining variation in use of financial services across and within countries // Brookings Papers on Economic Activity. 2013. Vol. 2013, No. 1. P. 279-340. DOI: 10.1353/eca.2013.0002.

3. Zins A., Weill L. The determinants of financial inclusion in Africa // Review of Development Finance. 2016. Vol. 6, No. 1. P. 46-57. DOI: 10.1016/j.rdf.2016.05.001.

4. Allen F., Demirguc-Kunt A., Klapper L., Martinez Peria M.S. The foundations of financial inclusion: Understanding ownership and use of formal accounts // Journal of Financial Intermediation. 2016. Vol. 27. P. 1-30. DOI: 10.1016/j.jfi.2015.12.003.

5. Lusardi A., Mitchell O.S. The economic importance of financial literacy: Theory and evidence // Journal of Economic Literature. 2014. Vol. 52, No. 1. P. 5-44. DOI: 10.1257/jel.52.1.5.

6. Cole S., Sampson T., Zia B. Prices or knowledge? What drives demand for financial services in emerging markets? // The Journal of Finance. 2011. Vol. 66, No. 6. P. 19331967. DOI: 10.1111/j.1540-6261.2011.01696.x.

7. Svyatov S.A., Bendyuk D.Yu. Financial inclusion: international experience and Kazakhstan case // Central Asian Economic Review. 2023. No. 2. P. 140-152. DOI: 10.52821/2789-4401-2023-2-140-152.

8. Klapper L., Singer D., Starita L., Norris A. The Global Findex Database 2025: Connectivity and financial inclusion in the digital economy. Washington, DC: World Bank, 2025. DOI: 10.1596/978-1-4648-2204-9.

9. Demirgüç-Kunt A., Klapper L., Singer D., Ansar S. The Global Findex Database 2021: Financial inclusion, digital payments, and resilience in the age of COVID-19. Washington, DC: World Bank, 2022. DOI: 10.1596/978-1-4648-1897-4.

10. World Bank. Global Findex: Methodology [Electronic resource]. URL: https://www.worldbank.org/en/publication/globalfindex/methodology (accessed: 28.09.2026).

UDC 620.9:004.9

Wei Wei,
Mynzhanova Gulzhakan,
Niyetalina Gaukhar,
Farabi Business School
Al-Farabi Kazakh National University
(Almaty, Kazakhstan)

DIGITAL–PHYSICAL INTEGRATION AND ECONOMIC PERFORMANCE IN INDUSTRIAL INTEGRATED ENERGY SYSTEMS: A SCENARIO-BASED ANALYSIS OF SMART GRID CONTROL, BATTERY ENERGY STORAGE, AND DIGITAL TWINS

Abstract: To identify the independent and synergistic effects of digital capabilities and physical flexibility on the economic performance of industrial integrated energy systems, this study develops a representative electricity–heat–gas integrated energy system for Kazakhstan's oil and gas processing sector. Annual electricity and useful-heat demands are fixed at 100 GWh and 150 GWhth, respectively. On a common physical infrastructure, eight 2^3 technology scenarios are constructed using smart-grid control (SG), a 4 MW/16 MWh battery energy storage system (BESS), and a digital twin (DT). The analysis combines 8,760-hour chronological dispatch, a 20-year incremental cash-flow assessment, Shapley value decomposition, sensitivity analysis, and Latin hypercube sampling to compare operating costs, peak grid purchases, renewable-energy curtailment, net present value (NPV), and marginal technology contributions. The results show that full integration under S7 reduces annual operating costs from USD 4.974 million to USD 4.189 million, generates annual savings of USD 785.0 thousand, and achieves the lowest grid-purchase peak and renewable-energy curtailment. However, its NPV of USD 2.251 million is lower than the USD 2.508 million achieved by S5, which combines smart-grid control and a digital twin. The annual Shapley contributions of SG, BESS, and DT are USD 281.0 thousand, USD 222.6 thousand, and USD 281.4 thousand, respectively, while synergy accounts for 25.2% of the total value of S7. Under the NPV criterion, the contribution of storage is negative, indicating that BESS can improve system operation without necessarily covering its capital and replacement costs. Sensitivity analysis shows that S7 outperforms S5 when battery costs decline by approximately 8.2%, electricity-price volatility increases by approximately 18.7%, renewable-energy scale increases by approximately 30.7%, or the real discount rate falls below approximately 6.7%. Accordingly, the study proposes a staged investment pathway in which forecasting and control capabilities are established first, followed by storage deployment once the relevant threshold conditions are met. The results represent a scenario simulation of a representative system under transparent assumptions rather than the observed performance of a specific firm.

Keywords: integrated energy system; smart grid; battery energy storage; digital twin; Shapley value; techno-economic assessment

1. Introduction

Integrated energy systems (IESs) coordinate electricity, heat, natural gas, renewable energy, storage, and flexible loads, thereby providing an important pathway for energy-intensive industries to reduce costs, improve reliability, and advance low-carbon transformation. As wind

power, photovoltaics, combined heat and power (CHP), and multiple load categories are connected simultaneously, system economic performance increasingly depends on hourly supply–demand matching, equipment condition, and forecast quality rather than installed capacity alone. Smart-grid control provides metering, communication, demand response, and equipment coordination; battery energy storage provides intertemporal energy shifting and rapid reserve; and digital twins connect operating data, equipment models, and forecasting algorithms to support load forecasting, fault identification, and predictive maintenance. These three technologies represent information connectivity, physical flexibility, and forward-looking decision capability, respectively, and their joint value may exceed the simple sum of their stand-alone benefits.

Global battery-storage costs have declined substantially. IRENA reports that the total installed cost of utility-scale BESS fell from USD 2,571/kWh in 2010 to USD 192/kWh in 2024; the IEA further reports that global additions of battery-storage capacity reached 108 GW in 2025, approximately 40% higher than in 2024. Nevertheless, cost reductions do not imply that all industrial users can achieve a positive NPV. Storage revenues remain constrained by peak–off-peak price spreads, maximum-demand charges, curtailment levels, degradation, and replacement costs. Compared with storage, smart control and digital twins can simultaneously affect electricity purchases, fuel consumption, equipment efficiency, forecast deviations, and unplanned outages, but their economic benefits must also cover software, sensors, data governance, and maintenance expenditures.

Kazakhstan provides a representative context for examining this issue. In 2024, the country generated 117.915 TWh of electricity and consumed 119.996 TWh, while renewable generation was approximately 7.56 TWh, or 6.4% of total generation. The national energy balance shows that oil and petroleum products, natural gas, heat, and electricity accounted for 32.9%, 21.1%, 16.9%, and 16.4% of final energy consumption, respectively. Continuous processing industries such as oil and gas processing must therefore improve energy utilization and digital operations while maintaining electricity and heat supply and production continuity.

The existing literature has three limitations. First, smart grids, storage, and digital twins are usually evaluated separately, with limited use of complete three-factor scenarios. Second, short-term operating-cost optimisation is rarely translated into NPV, internal rate of return (IRR), and investment sequencing. Third, the value of technology combinations is often not allocated fairly, making it difficult to distinguish stand-alone technology values from synergistic value. This study develops eight strictly comparable scenarios for a representative industrial IES and links hourly operation, life-cycle investment assessment, Shapley decomposition, and uncertainty analysis. It focuses on four questions: through which mechanisms do the three technologies create value; does joint deployment generate synergy; is the operating optimum also the financial optimum; and what staged investment sequence should firms adopt?

2. Literature Review and Research Positioning

Research on smart grids and integrated demand response shows that advanced metering, dynamic pricing, and cross-energy coordination can transform part of an industrial load into a dispatchable resource. Jiang et al. (2024) systematically reviewed load reduction and shifting, energy substitution, contractual response, and market interaction mechanisms; Wu et al. (2024) further connected capacity configuration, demand response, and predictive control. However, these studies generally focus on representative days or short-term operating costs and pay insufficient attention to the life-cycle returns of digital infrastructure.

Energy-storage research has gradually expanded from stand-alone arbitrage to value stacking that includes peak shaving, renewable-energy integration, reserve provision, and ancillary services. Goulart et al. (2024) emphasised the importance of function stacking for the economics of customer-side BESS, while Koubar et al. (2025) showed that market prices and service portfolios can substantially change storage payback periods. A techno-economic assessment of storage must incorporate charging costs, efficiency, degradation, calendar life, and mid-life replacement; otherwise, investment value may be overstated. Existing studies still rarely compare the marginal value of storage across different levels of digital maturity.

Digital-twin research has moved from equipment visualisation towards forecasting, scenario testing, and closed-loop optimisation. do Amaral et al. (2023) noted that energy digital twins can span production, storage, transmission, distribution, and consumption. Aguilera et al. (2024) combined online-calibrated models, demand forecasting, and MILP-based scheduling, demonstrating that economic benefits depend on forecast accuracy and equipment degradation. Nevertheless, enterprise-level assessments still lack a unified treatment of digital-twin capital costs, maintenance expenditure, and avoided outage losses.

This study positions SG, BESS, and DT within the same representative industrial IES, uses a 2^3 full-factorial scenario design to identify stand-alone and combined values, and applies the Shapley value to allocate average marginal contributions. Unlike causal empirical research, this is a techno-economic scenario analysis based on transparent parameters. Its contribution is to identify mechanisms, investment boundaries, and technology sequencing rather than estimate an average treatment effect for all firms in Kazakhstan.

3. Research Design and Methods

3.1. System Boundary, Data, and Core Assumptions

The study object is a representative industrial IES for Kazakhstan's oil and gas processing sector. The base year is 2024, the project operates for 20 years, and all monetary values are expressed in constant 2024 USD. Annual electricity demand is fixed at 100 GWh and annual useful-heat demand at 150 GWhth. The model uses standardised industrial hourly load and weather series and does not simulate changes in product output. The system includes the public grid, gas-fired CHP, a gas boiler, waste-heat recovery, 12 MW of photovoltaic capacity, 12 MW of wind capacity, and an optional 4 MW/16 MWh BESS. Net electricity exports and external heat sales are not permitted; thermal storage is excluded; and carbon pricing, green certificates, firm-specific subsidies, and product revenues are not included.

National electricity generation, consumption, and energy-structure data are drawn from official Kazakhstan statistics and KEGOC. Technology-cost ranges are based on DOE, IRENA, and peer-reviewed studies. Because firm-level contracts and SCADA data are unavailable, electricity prices, natural-gas prices, capacity charges, forecast deviations, and outage losses are represented by reproducible baseline scenario parameters.

Table 1. Core parameters and evaluation boundary of the representative system

Parameter	Baseline value	Unit	Description
Annual electricity/useful-heat demand	100 / 150	GWh / GWhth	Fixed in all scenarios
PV/wind capacity	12 / 12	MW	Capacity factors: 18% / 35%
CHP electricity/heat capacity	5.09 / 6.50	MW / MWth	Electrical efficiency: 36%; thermal efficiency: 46%
Boiler/waste-heat-recovery capacity	18.0 / 2.4	MWth	Boiler efficiency: 90%
BESS	4 / 16	MW / MWh	Activated only in S2, S4, S6, and S7
Average grid energy price	48	USD/MWh	Calibrated scenario parameter
Maximum-demand/capacity charge	30,000	USD/(MW·year)	Calibrated scenario parameter
Natural-gas price	18	USD/MWhLHV	Calibrated scenario parameter
Baseline imbalance/reliability cost	270 / 220	thousand USD/year	Expected annual loss parameters
Project life/real discount rate	20 / 8%	years / %	General inflation excluded

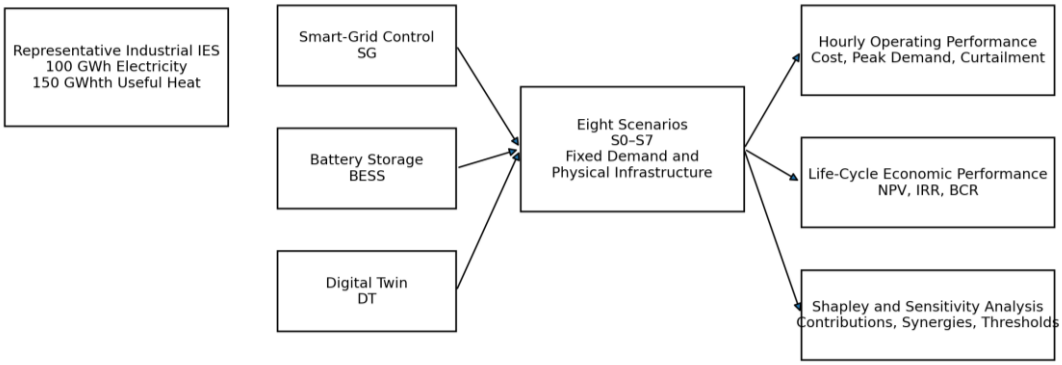


Figure 1. Research framework and 2^3 technology scenarios

3.2. Technology Scenarios

All scenarios use the same demand, weather, basic equipment capacities, and price series; only the availability of SG, BESS, and DT changes. S0 is not an unoptimised legacy plant but an "optimised physical baseline" without the three additional technologies, ensuring that all scenarios are comparable under the same dispatch boundary. SG enables intraday demand response, peak management, and real-time coordination; BESS provides charging, discharging, peak shaving, and reserve; and DT affects operation by reducing forecast errors, improving equipment availability and efficiency, and reducing unplanned outages.

Table 2. Eight technology scenarios for SG, BESS, and DT

Scenario	SG	BESS	DT	Principal capability
S0	0	0	0	No additional digital–physical technologies
S1	1	0	0	Demand response and coordinated control
S2	0	1	0	Stand-alone storage dispatch
S3	0	0	1	Forecasting, diagnosis, and maintenance optimisation
S4	1	1	0	Control plus physical flexibility
S5	1	0	1	Forecasting plus control
S6	0	1	1	Forecasting plus storage
S7	1	1	1	Forecast–control–execution closed loop

3.3. Hourly Operation and Economic Assessment

The model minimises annual operating costs, including electricity purchases, natural gas, operation and maintenance of basic equipment, demand-response compensation, storage degradation, digital-platform operation and maintenance, forecast-imbalance costs, and reliability losses. Electricity and heat balances are satisfied in every hour, subject to CHP and boiler output limits, renewable-energy availability, BESS state-of-charge constraints, mutually exclusive charging and discharging, and energy neutrality for demand response.

$$\min C_s^{op} = \sum_t \left(C_t^{grid} + C_t^{gas} + C_t^{OM} + C_t^{DR} + C_t^{deg} + C_t^{imb} + C_t^{out} \right) + C_s^{fix} \quad (1)$$

$$P_t^{grid} + P_t^{CHP} + P_t^{PV} + P_t^{WT} + P_t^{dis} = L_t^e + D_t^{up} - D_t^{down} + P_t^{ch} \quad (2)$$

$$Q_t^{CHP} + Q_t^{boiler} + Q_t^{WHR} = L_t^h \quad (3)$$

$$E_t = (1-\lambda)E_{t-1} + \eta_{ch} P_t^{ch} - P_t^{dis}/\eta_{dis} \quad (4)$$

DT does not enter the objective function through a uniform “energy-saving rate”; instead, it modifies forecast errors, equipment availability, and efficiency parameters. In the baseline calibration, DT increases CHP and boiler efficiency by 1.5% and 1.0%, respectively, on a relative basis and reduces expected imbalance and reliability losses. SG and BESS absorb temporal deviations through flexible-load and state-of-charge constraints. These improvement rates are varied in the sensitivity analysis.

Annual cost savings are defined as $AS_s = C_0^{op} - C_s^{op}$. The investment assessment uses 20-year incremental cash flows and includes initial investment, a digital-platform upgrade in Year 10, battery replacement in Year 12, and terminal residual value.

$$NPV_s = -I_{s,0} + \sum_{y=1}^{20} \frac{AS_s - I_{s,y}^{rep}}{(1+r)^y} + \frac{SV_s}{(1+r)^{20}} \quad (5)$$

Table 3. Scenario cash-flow assumptions (million constant 2024 USD)

Scenario	Initial investment	Annual net savings	Mid-life upgrade/replacement	Residual value	Principal assets
S1	1.05	0.2046	Year 10: 0.15	0.030	SG
S2	3.20	0.1533	Year 12: 1.12	0.320	BESS
S3	1.30	0.2294	Year 10: 0.20	0.040	DT
S4	3.85	0.4450	Year 10: 0.12; Year 12: 0.98	0.121	SG+BESS
S5	2.15	0.4865	Year 10: 0.28	0.050	SG+DT
S6	4.25	0.4210	Year 10: 0.18; Year 12: 1.02	0.121	BESS+DT
S7	5.00	0.7850	Year 10: 0.28; Year 12: 0.92	0.180	SG+BESS+DT

3.4. Shapley Decomposition and Uncertainty Analysis

SG, BESS, and DT are treated as players in a cooperative game, and the annual savings or incremental NPV of each scenario relative to S0 is used as the coalition value. The Shapley value of technology i is the weighted average of its marginal contribution when added to every possible coalition:

$$\phi_i(v) = \sum_{A \subseteq N \setminus \{i\}} \frac{|A|!\,(n-|A|-1)!}{n!}[v(A \cup \{i\}) - v(A)] \quad (6)$$

The one-factor analysis varies battery cost, electricity-price volatility, renewable-energy scale, load flexibility, DT cost, outage loss, and the discount rate. The two-dimensional scenarios focus on battery cost–electricity-price volatility, outage loss–DT cost, and load flexibility–electricity-price volatility. Joint uncertainty is evaluated using 300 Latin hypercube samples drawn from independent triangular distributions.

4. Results

4.1. Baseline Operating Performance

S0 satisfies annual electricity demand of 100 GWh and useful-heat demand of 150 GWhth without electricity or heat load shedding. The 12 MW PV and 12 MW wind installations provide 55.71 GWh of available generation, of which 55.28 GWh is utilised and 0.432 GWh is curtailed. CHP supplies 36.32 GWh of electricity, while 8.399 GWh is purchased from the public grid. Heat demand is supplied by 19.34 GWhth of waste-heat recovery, 46.41 GWhth of CHP heat, and 84.25 GWhth of gas-boiler heat.

The annual operating cost of S0 is USD 4.974 million. Natural-gas purchases account for USD 3.501 million, or 70.4% of the total. Grid-energy and capacity charges total USD 0.585 million, operation and maintenance of basic equipment amounts to USD 0.398 million, and imbalance and reliability losses total USD 0.490 million. The maximum grid-purchase power is 6.059 MW. This cost structure indicates that the value of digital technologies arises not only from electricity savings but also from fuel efficiency, reduced forecast deviations, and equipment reliability.

4.2. Operating and Investment Performance of the Technology Scenarios

Table 4. Core operating and investment results for S0–S7 (cost and NPV: million USD; savings: thousand USD/year; peak: MW; curtailment and natural gas: GWh)

Scenario	Annual cost	Annual savings	Grid-purchase peak	Curtailment	Natural gas	NPV	Investment assessment
S0	4.974	0.0	6.059	0.432	194.50	0.000	Baseline
S1	4.769	204.6	5.912	0.253	193.52	0.896	Marginally viable
S2	4.821	153.3	3.374	0.108	194.50	−2.071	Not viable under baseline conditions
S3	4.745	229.4	6.059	0.432	192.08	0.868	Marginally viable
S4	4.529	445.0	2.845	0.020	193.18	0.100	Marginally viable
S5	4.487	486.5	5.802	0.185	190.92	2.508	Strongly viable
S6	4.553	421.0	3.102	0.041	192.08	−0.579	Not viable under baseline conditions
S7	4.189	785.0	2.520	0.005	189.75	2.251	Strongly viable

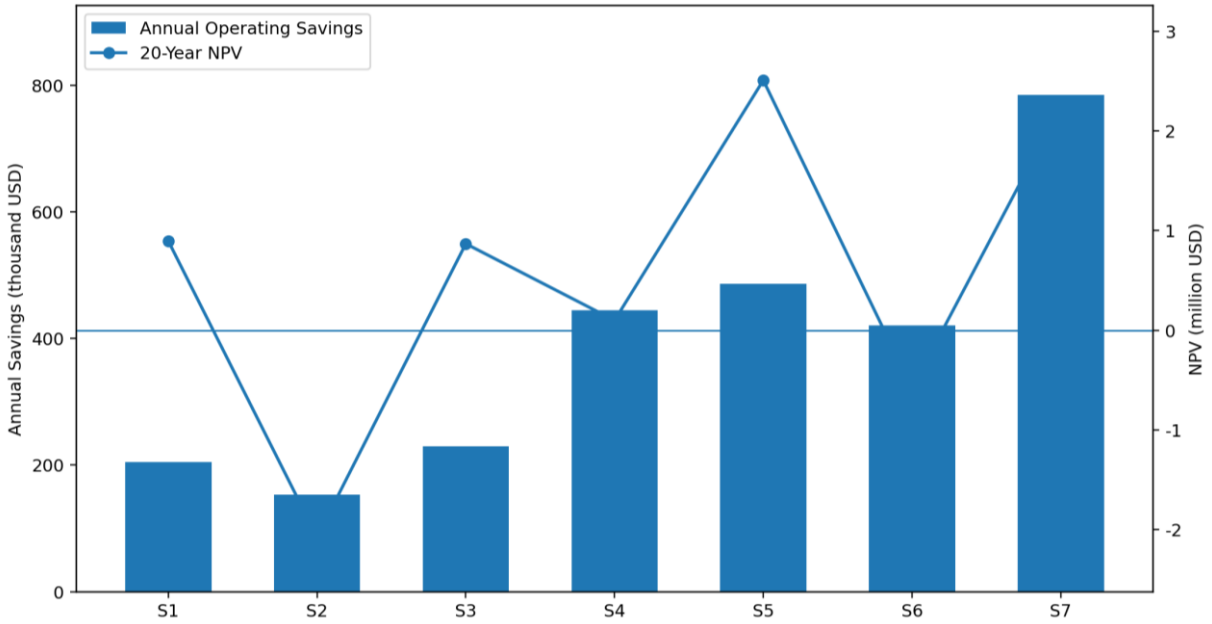


Figure 2. Annual operating value and life-cycle investment value of S1–S7

Among the stand-alone technologies, S3 produces the largest annual savings (USD 229.4 thousand), followed by S1 (USD 204.6 thousand) and S2 (USD 153.3 thousand). S1 mainly reduces imbalance and management costs, while S3 primarily lowers fuel use, outage losses, and maintenance costs. S2 reduces the grid-purchase peak from 6.059 MW to 3.374 MW and

lowers curtailment, but its stand-alone NPV is −USD 2.071 million. This result shows that physical flexibility has operating value but remains constrained by battery capital and replacement costs.

All pairwise combinations generate positive operating synergies. S4 has the strongest pairwise interaction benefit, USD 87.1 thousand/year, but its NPV is only USD 0.100 million, indicating a narrow safety margin. S5 generates annual savings of USD 486.5 thousand and an NPV of USD 2.508 million, making it the financially most valuable option under baseline conditions. S6 improves peak demand, forecasting, and reliability, but its NPV remains −USD 0.579 million.

S7 achieves the lowest annual operating cost, USD 4.189 million, which is 15.8% below S0. It reduces the maximum grid-purchase power to 2.520 MW and curtailment to 0.005 GWh and generates annual savings of USD 785.0 thousand. Its NPV is USD 2.251 million, its IRR is 13.8%, and its BCR is 1.41, indicating that the integrated project is economically viable. However, its NPV remains below that of S5. The incremental NPV of adding storage to S5 to form S7 is approximately −USD 0.257 million, demonstrating that the operating optimum is not necessarily the optimal capital-allocation decision.

Table 5. Twenty-year financial performance of the technology scenarios

Scenario	Initial investment	NPV	IRR	BCR	Discounted payback period	Principal conclusion
S1	1.05	0.896	18.3%	1.80	6.87 years	Low capital intensity and relatively rapid payback
S2	3.20	−2.071	−2.9%	0.43	Not recovered within 20 years	Stand-alone storage is not viable
S3	1.30	0.868	16.2%	1.62	7.85 years	Substantial reliability value
S4	3.85	0.100	8.4%	1.02	19.17 years	Near break-even
S5	2.15	2.508	21.8%	2.10	5.68 years	Highest NPV
S6	4.25	−0.579	6.0%	0.88	Not recovered within 20 years	Still constrained by storage costs
S7	5.00	2.251	13.8%	1.41	9.41 years	Operating optimum and overall viability

4.3. Shapley Value and Synergy

The Shapley decomposition of annual coalition value satisfies the efficiency property: the contributions of smart-grid control, battery storage, and the digital twin are USD 281.0 thousand/year, USD 222.6 thousand/year, and USD 281.4 thousand/year, respectively, summing to USD 785.0 thousand/year. The three stand-alone values total USD 587.3 thousand/year, while synergy is USD 197.7 thousand/year, accounting for 25.2% of the total value of S7. The SG–BESS, SG–DT, and BESS–DT synergies are USD 87.1 thousand/year, USD 52.5 thousand/year, and USD 38.3 thousand/year, respectively, and the higher-order three-way synergy is USD 19.8 thousand/year.

Under the NPV criterion, the Shapley contributions of SG, BESS, and DT are USD 1.877 million, −USD 1.150 million, and USD 1.524 million, respectively, summing to USD 2.251 million. The positive annual contribution but negative NPV contribution of storage reveals the coexistence of operating flexibility and a substantial capital burden. SG and DT make positive contributions under both the operating-value and investment-value criteria.

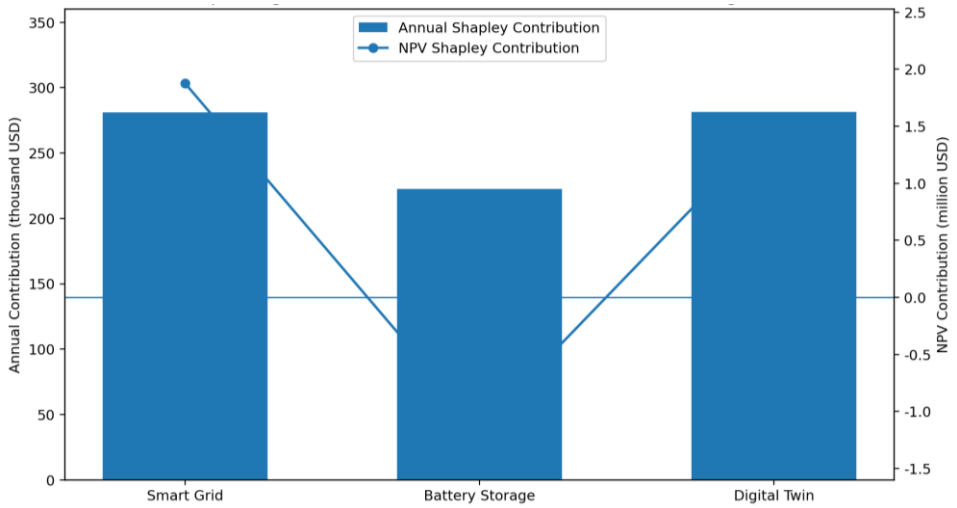


Figure 3. Annual operating contributions and life-cycle investment contributions of the three technologies

Table 6. Shapley contributions and synergistic value (annual value: thousand USD/year; NPV: million USD)

Technology/synergy	Annual value	Share of S7 value	NPV Shapley contribution	Interpretation
Smart-grid control	281.0	35.8%	1.877	High capital efficiency of control and demand response
Battery energy storage	222.6	28.4%	−1.150	Positive operating value but high capital cost
Digital twin	281.4	35.8%	1.524	Forecasting, efficiency, and reliability benefits

Technology/synergy	Annual value	Share of S7 value	NPV Shapley contribution	Interpretation
Total synergy	197.7	25.2%	—	Digital capabilities increase the value of physical flexibility

4.4. Sensitivity and Uncertainty

Under baseline conditions, S5 has the highest NPV, whereas S7 has the lowest operating cost. Threshold analysis shows that the NPV of S7 exceeds that of S5 when battery investment and replacement costs decline by approximately 8.2%, electricity-price volatility increases by approximately 18.7%, renewable-energy scale increases by approximately 30.7%, or the real discount rate falls below approximately 6.7%. By contrast, stand-alone storage under S2 requires a battery-cost reduction of approximately 57.9% to achieve an NPV of zero, indicating that digital control and forecasting capabilities substantially lower the economic threshold for storage.

Across 300 Latin hypercube samples, the probabilities of positive NPV are 100.0% for S5 and 99.7% for S7. S5 has the highest NPV in 57.3% of the samples, while S7 ranks first in 42.7%. The probabilities of positive NPV are 67.3% for S4 and 29.3% for S6, whereas S2 does not achieve a positive NPV in any sample. Joint uncertainty therefore supports a strategy of prioritising digital infrastructure and upgrading to storage conditionally.

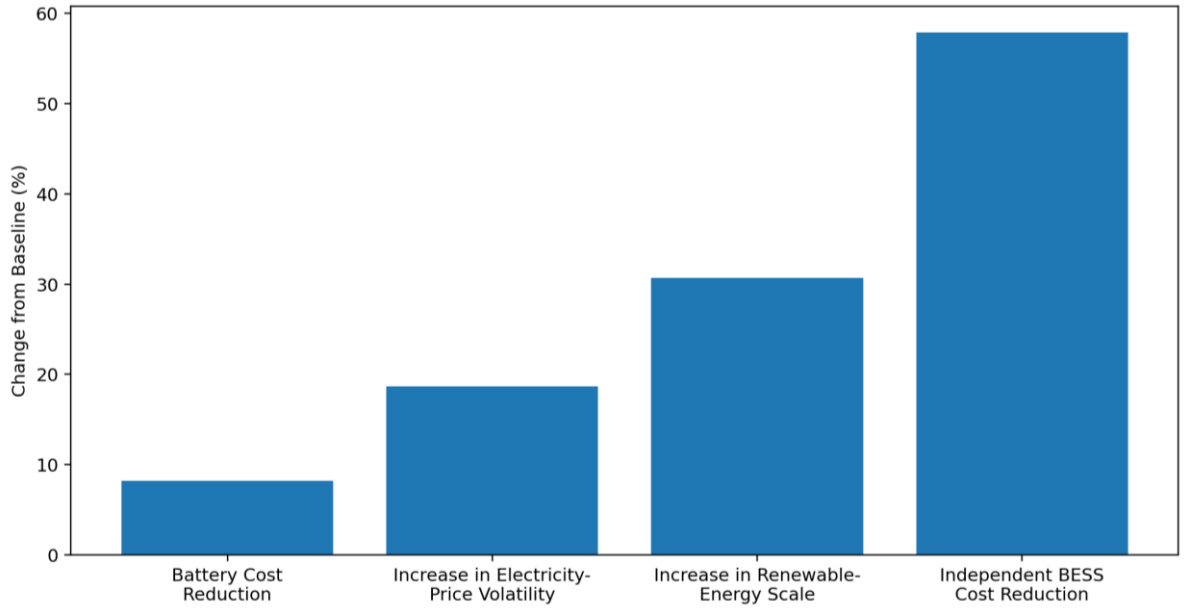


Figure 4. Key thresholds that change the investment ranking

Table 7. Key economic thresholds and investment implications

Decision question	Threshold condition	Baseline interpretation	Managerial implication
S7 exceeds S5	Battery cost decreases by 8.2%	Incremental NPV of storage becomes positive	Continuously update battery quotations and replacement costs
S7 exceeds S5	Electricity-price volatility increases by 18.7%	Arbitrage, peak-shaving, and balancing value increase	Monitor contractual price spreads and capacity charges
S7 exceeds S5	Renewable-energy scale increases by 30.7%	Temporal mismatch and curtailment value increase	Coordinate storage planning with additional wind and PV
S7 exceeds S5	Discount rate falls below 6.7%	Present value of long-term savings increases	Pursue low-cost green financing
S2 stand-alone BESS reaches NPV = 0	Battery cost decreases by 57.9%	Stand-alone arbitrage cannot cover capital costs	Avoid procuring storage independently of digital infrastructure

5. Discussion

5.1. Why the Operating Optimum Differs from the Investment Optimum

The fact that S7 has the best operating performance while S5 has the highest NPV is the central finding of this study. Hourly dispatch analysis asks how existing assets can be operated at minimum cost, whereas NPV analysis asks whether additional assets can cover initial investment, mid-life replacement, and the opportunity cost of capital. Compared with S5, S7 generates approximately USD 298.5 thousand in additional annual savings but requires USD 2.85 million in additional initial investment and battery replacement in Year 12. At an 8% discount rate, these incremental savings do not produce a positive incremental NPV.

5.2. Why Digital Infrastructure Increases the Value of Storage

The substantially negative NPV of stand-alone storage under S2, the near-break-even result for S4, and the overall viability of S7 show that storage value depends on system observability, predictability, and controllability. SG coordinates price signals and loads, allowing the battery to operate during high-value periods. DT reduces forecast deviations and unnecessary reserve occupancy, enabling more effective use of limited power and energy capacity. The SG–BESS synergy is the strongest pairwise synergy and supports an investment sequence in which digital infrastructure is established before physical flexibility is deployed.

5.3. Theoretical and Methodological Contributions

This study integrates digital capability and physical flexibility within a single techno-economic framework. The 2^3 scenario design avoids the attribution problem created when only a baseline and a fully integrated scenario are compared. The Shapley decomposition further shows that approximately one quarter of the annual value of S7 arises from synergy. The dual decomposition of annual value and NPV also reveals that a technology can make a substantial operating contribution while generating a negative investment contribution because of capital costs. This result extends the concept of "flexibility value" in multi-energy-system research by incorporating both functional value and capital efficiency.

5.4. External Applicability and Research Limitations

The study uses a representative system and standardised hourly series and is therefore neither a firm-level empirical analysis nor a causal identification exercise. Electricity prices, outage losses, and digital-technology improvement rates are scenario parameters. BESS capacity is fixed at 4 MW/16 MWh rather than optimised endogenously. Net electricity exports, ancillary-service revenues, thermal storage, carbon pricing, taxation, and firm-specific subsidies are excluded. The sensitivity analysis uses independent triangular distributions and cannot be interpreted as an objective probability forecast of future markets.

6. Conclusions and Managerial Implications

This study compares the stand-alone, pairwise, and fully integrated values of smart-grid control, BESS, and digital twins in a representative industrial IES. The results show that the S5 combination of smart-grid control and a digital twin has the highest NPV under baseline conditions, while S7 full integration achieves the lowest annual operating cost and the best performance in peak demand, curtailment, and reliability. Storage consistently creates operating flexibility, but its stand-alone and average NPV contributions remain constrained by capital and replacement costs. Digital–physical synergy accounts for 25.2% of the annual value of full integration, indicating that stand-alone asset assessments underestimate the value of joint deployment.

Firms should adopt a staged pathway. First, they should establish unified metering, data governance, and an energy baseline. Second, smart-grid control and an energy management system should be deployed to enable demand response and multi-equipment coordination. Third, digital twins should be applied to load forecasting, equipment-condition identification, and predictive maintenance. Fourth, battery costs, electricity-price volatility, renewable-energy scale, and financing costs should be monitored continuously, and storage should be installed or expanded only after its incremental NPV becomes positive. Operating departments should use peak demand, curtailment, imbalance, and reliability KPIs to assess technical functionality, whereas investment departments should use incremental NPV, IRR, and BCR to make capital-budgeting decisions.

For Kazakhstan's oil and gas processing sector, the primary question in digital transformation is not whether all technologies should be adopted simultaneously, but how information and decision capabilities can first be established at the lowest capital cost and physical flexibility subsequently added according to observable market and operating thresholds. This pathway balances near-term economic returns with the option to upgrade towards a complete digital–physical closed loop in the future.

References:

1. Aguilera, J. J., Padullés, R., Meesenburg, W., Markussen, W. B., Zühlsdorf, B., & Elmegaard, B. (2024). Operation optimization in large-scale heat pump systems: A scheduling framework integrating digital twin modelling, demand forecasting, and MILP. Applied Energy, 376, 124259.

2. Bâra, A., & Oprea, S.-V. (2024). Enabling coordination in energy communities: A digital twin model. Energy Policy, 184, 113910.

3. Chatzigeorgiou, N. G., Theocharides, S., Makrides, G., & Georghiou, G. E. (2024). A review on battery energy storage systems: Applications, developments, and research trends of hybrid installations in the end-user sector. Journal of Energy Storage, 86, 111192.

4. do Amaral, J. V. S., dos Santos, C. H., Montevechi, J. A. B., & de Queiroz, A. R. (2023). Energy digital twin applications: A review. Renewable and Sustainable Energy Reviews, 188, 113891.

5. Goulart, V. V., Rodrigues Neto, J. A., & Rüther, R. (2024). Method of techno-economic analysis of battery energy storage system function-stacking for medium-voltage connected consumers. Journal of Energy Storage, 83, 110754.

6. Jiang, M., Xu, Z., Zhu, H., Goh, H. H., Kurniawan, T. A., Liu, T., & Zhang, D. (2024). Integrated demand response modeling and optimization technologies supporting energy internet. Renewable and Sustainable Energy Reviews, 203, 114757.

7. Koubar, M., Lindberg, O., Lingfors, D., Huang, P., Berg, M., & Munkhammar, J. (2025). Techno-economical assessment of battery storage combined with large-scale photovoltaic power plants operating on energy and ancillary service markets. Applied Energy, 382, 125200.

8. Li, S., Zhou, Y., Wu, J., Pan, Y., Huang, Z., & Zhou, N. (2025). A digital twin of multiple energy hub systems with peer-to-peer energy sharing. Applied Energy, 380, 124908.

9. Wu, X., Yang, L., & Zheng, B. (2024). Joint capacity configuration and demand response optimization of integrated energy system considering economic and dynamic control performance. Energy, 301, 131723.

UDC 339.138

Xu Fumin,
Wei Wei,
Ruziyeva Elvira,
Moldabekova A.
Farabi Business School
Al-Farabi Kazakh National University
(Almaty, Kazakhstan)
Shandong College of Tourism and Hospitality
(Jinan, China)
Shandong Province Tourism Development Research Center
(Jinan, China)

FROM CULTURAL RESOURCES TO GLOBAL CONSUMER VALUE: THE BRIDGE–IMC MODEL AND THE INTERNATIONALIZATION OF BLACK MYTH: WUKONG UNDER THE BELT AND ROAD INITIATIVE

Abstract. The global circulation of cultural products depends on more than the possession of distinctive heritage. Cultural resources must be integrated into a marketable experience, distributed through digital systems and converted into value that foreign consumers can understand and participate in creating. This article develops the BRIDGE–IMC model through a theory-elaborating case study of Black Myth: Wukong. Evidence from 2020 to July 2026 is triangulated across producer announcements, platform records, technology-partner materials, independent academic studies and institutional reports. BRIDGE represents six connected dimensions: the Belt and Road institutional context (B), relationship building (R), integration of cultural resources (I), digital marketing capability (D), global market expansion (G), and consumer engagement (E). International marketing capability (IMC) provides the orchestration logic linking these dimensions. The case shows that Chinese mythology, historic architecture, music, visual art and advanced game technology were assembled into a coherent global consumer experience. Digital visibility and platform access supported rapid scale, while participation through play, reviews, livestreams and user-generated content converted cultural meaning into experiential, symbolic and social value. However, the concentration of early demand in East Asia demonstrates that global availability is not equivalent to balanced international diversification. The model proposes that consumer engagement creates a recursive learning loop that renews relationships, resource integration and digital routines. The contribution is a process explanation of how cultural resources become consumer value, brand competitiveness and international market performance. The framework also provides a diagnostic tool for China–Kazakhstan cultural-industry projects under the Belt and Road Initiative.

Keywords: BRIDGE–IMC model; cultural industries; international marketing capability; Black Myth: Wukong; consumer engagement; Belt and Road Initiative

Introduction

Cultural resources do not enter international markets by themselves. A classical narrative, architectural tradition, musical form or visual symbol may possess substantial cultural value and still generate little foreign-market value if it cannot be organized into a product that audiences

can access, understand and use. The strategic problem for cultural industries is therefore one of conversion: how can culturally specific resources become consumer value without losing the difference that makes them distinctive?

Digital games make this problem unusually visible. They combine software, narrative, visual art, music, performance, design and intellectual property in one interactive product. They also circulate through global platforms on which attention, reviews, livestreams and user-generated content affect discovery and demand. Recent research describes video games as an important site of cultural production rather than a marginal entertainment form (Yodovich et al., 2025). Their production depends on ecosystems that cross technology, media, heritage and creative sectors (Klimas et al., 2026). International success consequently requires more than high production quality. It requires a capability to coordinate relationships, cultural meanings, digital channels, market expansion and consumer participation.

Black Myth: Wukong provides a revealing case. Developed by Game Science and released in August 2024, the action role-playing game draws on Journey to the West and a dense field of Chinese mythological, architectural, musical and philosophical references. Guinness World Records reports that it sold 20 million units in its first month and exceeded 2.2 million concurrent Steam players on its first day (Guinness World Records, n.d.). By 30 July 2026, its Steam page displayed more than 875,000 reviews across languages and an ‘Overwhelmingly Positive’ overall evaluation (Steam, n.d.). These outcomes make the case theoretically useful, but the evidence also calls for restraint. Launch demand was concentrated in East Asia, and independent research finds that international players often valued technical and gameplay accessibility while missing parts of the cultural narrative (Gao et al., 2025). Global availability and global market diversification are not the same result.

Existing international marketing capability research explains how firms sense markets, build relationships, implement marketing strategies and reconfigure resources (Morgan, 2012; Morgan et al., 2018; Pfajfar, Mitręga, & Shoham, 2024a). Digital marketing research adds scalable communication, analytics and customer interaction (Homburg & Wielgos, 2022; Wielgos et al., 2021). Cultural-industry research explains symbolic value and the tension between authenticity and market accessibility (Morhart et al., 2015; Throsby, 2008). Yet these streams do not fully explain the process through which cultural resources move from organizational inputs to consumer value and international market performance.

This article develops the BRIDGE–IMC model to provide that process explanation. BRIDGE denotes the Belt and Road institutional context (B), relationship building (R), integration of cultural resources (I), digital marketing capability (D), global market expansion (G), and consumer engagement (E). IMC refers exclusively to international marketing capability. The model extends the prior hierarchical CIIMC framework by converting its macro, foundational, core and strategic layers into an observable case process. B is the macro context; D supplies foundational digital execution; R and I perform relational and cross-cultural conversion; G provides strategic expansion; and E turns market response into consumer value and renewal feedback.

Three research questions guide the study. First, how were Chinese cultural resources integrated and marketed in the internationalization of Black Myth: Wukong? Second, how do the BRIDGE dimensions interact to convert those resources into consumer value? Third, what can this case contribute to the design of international marketing capability for China–Kazakhstan cultural-industry cooperation under the Belt and Road Initiative (BRI)? The article contributes a mid-range process model and a set of testable propositions. It does not claim that

the game was an official BRI project; rather, it uses the case to identify firm- and ecosystem-level mechanisms that BRI cultural cooperation can enable.

Literature review

International marketing capability in digital cultural markets

Marketing capability is the repeatable organizational ability to deploy resources in order to understand and serve markets. Day's (1994) distinction among inside-out, outside-in and spanning processes remains useful: firms need operational routines, market-sensing processes and mechanisms that connect the two. In international settings, capability must additionally coordinate market knowledge, strategy formation and implementation across institutional and cultural boundaries (Morgan et al., 2018). Export performance depends not on possession of a strategy alone but on the ability to execute and adapt it.

Dynamic capability theory explains why the capability base must change. Sensing, seizing and reconfiguring allow organizations to respond to technological and market discontinuity (Teece, 2007). International dynamic marketing capabilities are particularly relevant when platforms, audience tastes and institutional conditions change simultaneously (Hoque et al., 2022; Pfajfar et al., 2024a, 2024b). The distinction between ordinary and dynamic capability matters in a cultural project: releasing content reliably is an operating capability; changing the narrative format, partner configuration or business model is a strategic renewal capability.

Digital marketing capability forms the operational infrastructure of this process. It combines content distribution, platform management, customer-data use, campaign experimentation and real-time response. Such capabilities have direct value relevance for firm performance and customer outcomes (Homburg & Wielgos, 2022; Wielgos et al., 2021), while the wider digital-marketing literature emphasizes the interaction of technology, content, consumer behavior and governance (Dwivedi et al., 2021). Their contribution is increasingly treated as part of a capability hierarchy rather than as a stand-alone tool set (Manis & Madhavaram, 2023). A recent review in emerging-economy settings similarly identifies marketing analytics, social media, customer engagement and organizational readiness as connected drivers of innovation and performance (Elkhoudary et al., 2026).

For cultural products, however, digital access is an incomplete outcome. A trailer can be seen globally while the narrative remains culturally opaque. The international marketing capability of a cultural enterprise must therefore link platform execution to cultural interpretation, relational legitimacy and innovation. This is the theoretical position from which BRIDGE–IMC begins. Firm-level cultural intelligence and social-media technology can facilitate this link, but only when they are institutionalized as repeatable marketing routines (Hazzam & Wilkins, 2022).

Cultural-resource integration, authenticity and consumer value

Cultural resources include stories, symbols, places, rituals, artistic practices, skills and intellectual property. Their value is not additive. A product does not become more meaningful simply by accumulating visible signs of heritage. Integration is required: resources must support a coherent narrative, experience and brand promise. In video games, the integration challenge is especially demanding because cultural meaning must work through mechanics, environment, sound, character design and player action.

Authenticity is central to the evaluation of this integration. Consumers may regard a cultural brand as authentic when it appears credible, internally consistent, culturally grounded and faithful to its claimed identity (Morhart et al., 2015). Cross-cultural evidence also shows

that perceived authenticity mediates the relationship between brand experience and stronger consumer-brand attachment, although effects vary across markets (Rodrigues et al., 2024). Authenticity should therefore not be treated as the absence of adaptation. It is an audience judgment about whether adaptation preserves a meaningful source identity.

The relevant output is consumer value. For an interactive cultural product, value can be functional (technical quality and playability), experiential (immersion and emotion), symbolic (identity, cultural meaning and status), and social (participation in communities and shared interpretation). Cultural value and commercial value can reinforce one another, but they can also diverge. A technically successful game may leave international players outside its deeper narrative, while a culturally respected work may fail to reach a scalable audience.

The latest reception research on Black Myth: Wukong demonstrates this tension. Gao et al. (2025) find that Chinese audiences emphasized cultural authenticity and identity, whereas many international players focused on gameplay, technology and accessibility. Earlier case analysis identified sensory, symbolic, translation, technical and playability factors as joint conditions of cross-cultural acceptance (Mao, 2024), while research on Chinese mobile games shows that ‘Chineseness’ is actively debated inside online communities (Li & Li, 2023). The findings suggest that integration must be evaluated twice: first within the product, and then in the meanings consumers actually construct.

Relationships, global expansion and consumer engagement

International cultural production is ecosystem-based. Developers require platforms, hardware and software partners, distributors, translators, creators, media, cultural institutions and audience communities. The relational view explains how complementary assets, knowledge sharing and governance can create value that no single organization controls (Dyer & Singh, 1998). Current video-game research likewise identifies cooperation, knowledge management and adaptive coordination as central to cross-industry projects (Klimas et al., 2026). Relationships are therefore not merely a promotional channel; they are productive infrastructure.

Global expansion adds market selection, localization, positioning, pricing, platform access and scale. Standardization and adaptation are not opposing solutions but contingent choices that must fit product and market conditions (Mandler et al., 2021). Digital platforms can make a cultural product immediately available across borders, but they also concentrate data and bargaining power. Platformization reorganizes cultural production around programmable infrastructures, visibility rules and continuous metrics (Nieborg & Poell, 2018).

Consumer engagement changes the direction of value creation. It includes cognitive attention, emotional investment and behavioral participation through use, reviews, recommendation, content creation and community activity. Social-media strategy turns audiences from recipients into participants (Li et al., 2021), while engagement research emphasizes the managerial importance of listening, interaction and user-generated content (Lim & Rasul, 2022). Cross-country evidence further shows that cultural conditions influence clicking, sharing and electronic word of mouth (Mattison Thompson & Brouthers, 2021). Engagement is consequently both an outcome of international marketing and a source of new market knowledge.

Digital content can build engagement, trust and value when it is useful, relevant and narratively coherent (Hollebeek & Macky, 2019). It can also amplify misunderstanding or controversy. BRIDGE–IMC therefore treats engagement as a recursive mechanism: consumers receive and interpret cultural value, but their responses also reshape relationships, integration choices and digital routines.

The BRI and the China–Kazakhstan relevance

The BRI is an institutional setting for connectivity and exchange rather than a substitute for firm capability. Research on BRI markets shows that cultural friction can affect export performance even when infrastructure and policy access improve (Li et al., 2019). Kazakhstan's BRI engagement also combines national development logic with international institutional imagination (Lee et al., 2021). The implication is that policy opportunity must be translated into relationships, products and audience value.

This translation is relevant to Kazakhstan's growing creative economy. Recent studies emphasize the role of universities, cities and cross-sector networks in the development of Kazakhstani creative industries (Kim & Comunian, 2024), while a Central Asian perspective links the creative economy to diversification, identity and regional cooperation (Auken et al., 2025). Official information reported 45,500 registered creative-industry entities in Kazakhstan in 2025, with strong concentration in Almaty and Astana and a large share of individual entrepreneurs (Government of the Republic of Kazakhstan, 2025). Smaller cultural firms need shared capabilities and relationships precisely because they cannot independently finance every internationalization function.

The literature thus leaves a process gap. International marketing explains capability; cultural studies explain meaning; platform research explains digital circulation; BRI research explains institutions. BRIDGE–IMC links these levels around a single conversion question: how does a cultural resource become consumer value and market performance?

Methodology

Research design and case selection

The study uses a theory-elaborating single-case design. A single case is appropriate when the phenomenon is revelatory, theoretically significant and embedded in a complex real-world context (Yin, 2018). Extreme or highly visible cases can expose mechanisms that are difficult to observe in average cases, provided the study does not generalize frequencies from one observation (Eisenhardt, 1989). Black Myth: Wukong was selected because it combines a culturally dense source, advanced digital production, rapid market scale, extensive public traces and evidence of both successful engagement and cross-cultural limitation.

The unit of analysis is the international market development of the game as a cultural product, not Game Science as an entire firm. The observation period begins with the first widely circulated gameplay reveal in 2020 and continues to 30 July 2026. The period captures pre-release attention formation, partner and platform preparation, launch, market response, cultural spillovers and post-launch reception.

The case is used to elaborate a process theory, not to test the causal effect of the BRI. No reviewed source identifies the game as an official BRI project. Accordingly, the B dimension is specified as the institutional context in which the case mechanisms can be transferred to cross-border cultural cooperation. This boundary avoids a common inferential error: assigning a policy label to a commercial outcome without evidence of a direct program relationship.

Evidence collection and analysis

The evidentiary archive was constructed from five source groups. Producer-controlled records supplied release and positioning information. Platform and market records supplied dated indicators of availability, reviews, concurrent use, sales and awards. Technology partner records documented complementary production and marketing relationships. Peer-reviewed studies provided independent analysis of cultural representation and reception. Government and

institutional reports provided evidence of cultural-tourism spillovers and the policy context. Table 1 summarizes the source logic.

Table 1 – Case evidence and its analytical role

Evidence group	Examples and period	Analytical use	Reliability and limitation
Producer-controlled records	Game Science product descriptions, trailers and official sales announcement, 2020–2025	Positioning, release sequence, cultural narrative and first-party performance claims	Direct evidence of intent and milestones; promotional framing requires triangulation
Platform and market records	Steam store/review page; Guinness World Records; The Game Awards, 2024–2026	Global availability, engagement scale, sales milestones and external recognition	Observable and dated; platform metrics are dynamic and do not reveal all regional sales
Technology and ecosystem records	NVIDIA partnership announcements and technical documentation, 2024	Relationship building, digital production capability and global co-marketing	Partner-verified; focuses on technology users and omits confidential arrangements
Independent research	Peer-reviewed studies of cross-cultural communication and audience reception, 2024–2026	Cultural integration, authenticity, accessibility and asymmetric reception	Provides methodological independence; early scholarship remains case-specific
Institutional and spillover records	Chinese government cultural-tourism reports; Kazakhstan creative-industry policy information, 2024–2025	Cultural spillovers and transferability to BRI cultural-industry development	Useful for policy outcomes; not evidence that the game was an official BRI project

Note – all sources are publicly accessible; dynamic platform values are reported with an observation date.

Analysis proceeded in three rounds. In the first round, evidence was coded in descriptive terms, including partner resource, cultural symbol, narrative integration, technical capability, platform reach, localization, review activity, audience interpretation and spillover. In the second round, these codes were grouped under relationship, integration, digital, global and engagement mechanisms. In the third round, the mechanisms were mapped onto the earlier hierarchical CIIMC logic and connected to consumer value, brand competitiveness and performance. The BRI context was added only at the institutional level.

Triangulation emphasized convergence and contradiction. A producer statement was not used alone when an external record was available. Dynamic metrics are reported with dates. Contradictory evidence—especially the difference between global sales visibility and the regional concentration of demand—was retained because it clarifies the boundary between

global availability and global diversification. The resulting propositions are analytical generalizations for later testing, not statistical estimates.

Results and discussion

Case overview: exceptional scale with an uneven global footprint

Black Myth: Wukong entered the market as an action role-playing game rooted in Chinese mythology and Journey to the West (Steam, n.d.). Its distinctiveness came from the combination of a globally recognizable interactive format with source-specific characters, environments, architecture, music and philosophical references. Technology was not separate from cultural presentation. NVIDIA documented cooperation with Game Science to incorporate full ray tracing and DLSS technology, linking production quality with access to an established global hardware and PC-gaming ecosystem (NVIDIA, 2024).

The launch outcomes were exceptional. Game Science announced sales of 10 million units across platforms by 23 August 2024 (Game Science, 2024), and Guinness World Records later certified 20 million units in the first month (Guinness World Records, n.d.). The game also won Best Action Game at The Game Awards 2024 (The Game Awards, 2024). These indicators show rapid conversion of attention into purchase and recognition.

Nevertheless, the case did not immediately produce a geographically balanced consumer base. Guinness reports that 93% of Steam reviews at launch were in simplified Chinese and 82% of launch-day traffic originated in East Asia (Guinness World Records, n.d.). Independent analysis later found asymmetric reception across Chinese and international audiences (Gao et al., 2025). The case is therefore more informative than a simple success story: it shows both the power and the unfinished work of international marketing capability.

B – Belt and Road context: enabling condition, not direct cause

The B component situates cultural internationalization within the connectivity, exchange and legitimacy created by the BRI. It may improve access to partners, events, digital infrastructure, policy attention and cross-border cultural programs. In China–Kazakhstan cooperation, these conditions can reduce the cost of initiating a project and help small firms find institutions and audiences.

The Black Myth case also establishes the limit of this reasoning. Its observed market performance is explained most directly by product integration, technology, platform reach, engagement and consumer response. The evidence does not support a direct BRI-performance claim. B therefore operates through organizational and ecosystem mechanisms; it changes the opportunity structure but does not create consumer value by administrative designation.

Proposition 1. The Belt and Road institutional context supports international cultural-market development indirectly through relationship formation, resource access and capability development; it has no sufficient direct effect on consumer value or market performance.

R – Relationship: mobilizing complementary assets

The case reveals several types of relationship. Technology cooperation linked Game Science with global hardware and software ecosystems. Digital storefronts and console channels connected the product to purchasing, delivery and review infrastructures. Streamers, reviewers and player communities acted as cultural intermediaries. After launch, tourism agencies and heritage sites connected the digital experience to physical places.

These relationships performed different tasks. Technical partners strengthened performance and global signaling; platforms provided reach and data; creators reduced search costs; players supplied interpretation and electronic word of mouth; cultural institutions

extended the value chain beyond the game. The relationship dimension is therefore a portfolio of complementary assets rather than a generic measure of networking.

Relationship quality also shapes cultural governance. A partner may broaden reach while simplifying cultural meaning, or provide technology while increasing dependence. Effective IMC requires clear roles, data rights, intellectual-property safeguards, attribution and mechanisms for joint learning.

Proposition 2. Relationship capability increases the quality of cultural-resource integration by providing complementary technology, market knowledge, legitimacy and distribution assets, subject to balanced governance and knowledge sharing.

I – Integration: converting heritage into an interactive value proposition

Integration is the core cultural conversion mechanism. Black Myth did not present Journey to the West as a digital museum. It reorganized narrative fragments, mythological figures, architecture, sculpture, landscape, music and combat traditions around player action. Real cultural locations were converted into explorable spaces, while a familiar action-RPG grammar reduced the initial learning cost for international players.

The integration was both material and semantic. Advanced graphics increased the visibility of architectural and environmental detail, but technical fidelity alone could not explain cultural value. Meaning emerged from the relationship among image, sound, challenge, character and narrative context. Gao et al. (2025) describe this as a hybridization of tradition and global game conventions. The case supports a complementary rather than substitutive view: global conventions become a carrier for source culture when they do not displace the source identity.

The physical spillover illustrates successful integration. Official reports identified 36 real-world locations represented in the game, including 27 in Shanxi, and linked the title to increased tourism interest (State Council Information Office of the People's Republic of China, 2024). A digital product thus reactivated value in architecture, destinations and heritage services. This is evidence of cultural-resource multiplication, not only game sales.

Proposition 3. Cultural-resource integration increases perceived authenticity, differentiation and experiential value when global product conventions complement rather than replace the source culture's narrative and symbolic coherence.

D – Digital: building reach, access and market intelligence

The digital dimension includes both production and marketing execution. High-fidelity development made cultural settings experientially credible, while trailers and platform pages allowed the product to be discovered, wish-listed, purchased and discussed across borders. Platform reviews then created a continuous market record. Digital capability connected content, distribution, analytics and response.

The launch illustrates the value of accumulated attention. Publicly circulated trailers created a long pre-release period in which audiences, creators and technology partners interpreted the product before purchase. After launch, more than 875,000 Steam reviews by July 2026 supplied social proof and a large body of market feedback (Steam, n.d.). Digital visibility thereby reduced uncertainty for new consumers and generated data for renewal.

Yet visibility is not cultural understanding. Digital systems can scale a misunderstanding as easily as a valued meaning. The D dimension becomes an international marketing capability only when analytics and content decisions are connected to integration and cross-cultural interpretation.

Proposition 4. Digital marketing capability increases scalable international access and market intelligence, but its effect on consumer value is mediated by cultural-resource integration and market-specific interpretation.

G – Global: separating access from diversification

Global capability comprises market selection, localization, positioning, platform coordination and the ability to scale without destroying product coherence. Black Myth achieved worldwide digital availability and rapid commercial scale. Its globally familiar genre and technical standard reduced category uncertainty, while its Chinese cultural specificity generated differentiation.

However, the regional concentration of launch traffic shows that a global channel does not guarantee global market depth. Three stages should be distinguished: global availability, when consumers can buy the product; global adoption, when meaningful demand emerges in several markets; and global diversification, when performance is not dominated by one home or regional base. The case clearly reached the first stage and demonstrated elements of the second, while the evidence is insufficient to claim fully balanced diversification.

This distinction is useful for BRI projects. A cultural producer should not evaluate internationalization solely by the number of countries covered by a platform. Market portfolios should be examined through non-home revenue, retention, review-language distribution, local creator participation and repeated demand.

Proposition 5. Digital platform reach positively influences global market expansion, but international market performance becomes more resilient only when global availability is converted into geographically diversified adoption through localization, positioning and relationship capability.

E – Engagement: co-creating meaning and renewal feedback

Engagement begins with play but extends to reviewing, livestreaming, recommending, explaining symbols, creating guides and visiting associated cultural sites. These activities create experiential, symbolic and social value. They also turn consumers into cultural intermediaries who translate the product for other consumers.

The case shows that engagement is not uniformly positive. International audiences may praise combat and visual quality while overlooking narrative depth; controversy or technical problems can also circulate rapidly. This variance is analytically useful. Reviews and community discussion reveal where cultural meaning has been received, adapted or lost. Engagement therefore mediates value conversion and supplies data for capability renewal.

The feedback path is essential. Consumer responses can change localization, explanatory content, platform priorities, partner selection and future product design. A one-off viral launch becomes a durable international capability only when engagement knowledge is retained and transferred across projects.

Proposition 6. Consumer engagement mediates the relationship between global market expansion and consumer value, while engagement feedback positively influences the renewal of relationship, integration and digital marketing capabilities.

Table 2 – BRIDGE–IMC dimensions, case manifestations and value-conversion roles

Dimension	Theoretical function	Manifestation in Black Myth: Wukong	Converted value	Characteristic risk
B – Belt and Road context	Macro-level connectivity and legitimacy	Not a project input documented by Game Science; used as the institutional extension context for cross-border cultural industries	Opportunity, access and policy coordination	Attributing firm success directly to policy
R – Relationship	Relational infrastructure and complementary assets	Developer-platform-hardware cooperation; distributor, creator, player and tourism-ecosystem links	Technical quality, reach, legitimacy and knowledge	Platform dependence or unbalanced appropriation
I – Integration	Recombination of cultural, creative and technological resources	Journey to the West, real architecture, music and myth integrated with an AAA action-RPG format	Authenticity, novelty and coherent experience	Decorative symbolism or cultural dilution
D – Digital	Scalable communication, analytics and platform execution	Viral trailers, digital storefronts, global content circulation, high-fidelity production and review data	Awareness, accessibility, data and conversion	Visibility without comprehension
G – Global	Market selection, localization, positioning and scaling	Worldwide PC/console launch and multilingual access, but geographically concentrated early demand	International reach and revenue diversification	Confusing global availability with global diversification
E – Engagement	Participation, meaning co-creation and market feedback	Play, livestreams, reviews, fan discussion, user-generated content and tourism interest	Experiential, symbolic and social value; learning	Misinterpretation, controversy or one-off attention

Note – IMC denotes international marketing capability, not integrated marketing communications.

The BRIDGE–IMC process model

Figure 1 integrates the findings. Cultural resources enter the model as potential value. The BRI context can facilitate exchange but remains outside the firm-level conversion chain. Relationships mobilize complementary assets; integration turns those assets into a coherent cultural proposition; digital capability provides access, distribution and data; global capability configures markets; engagement converts reception into value and learning. IMC orchestrates the sequence across foundational, core and strategic layers.

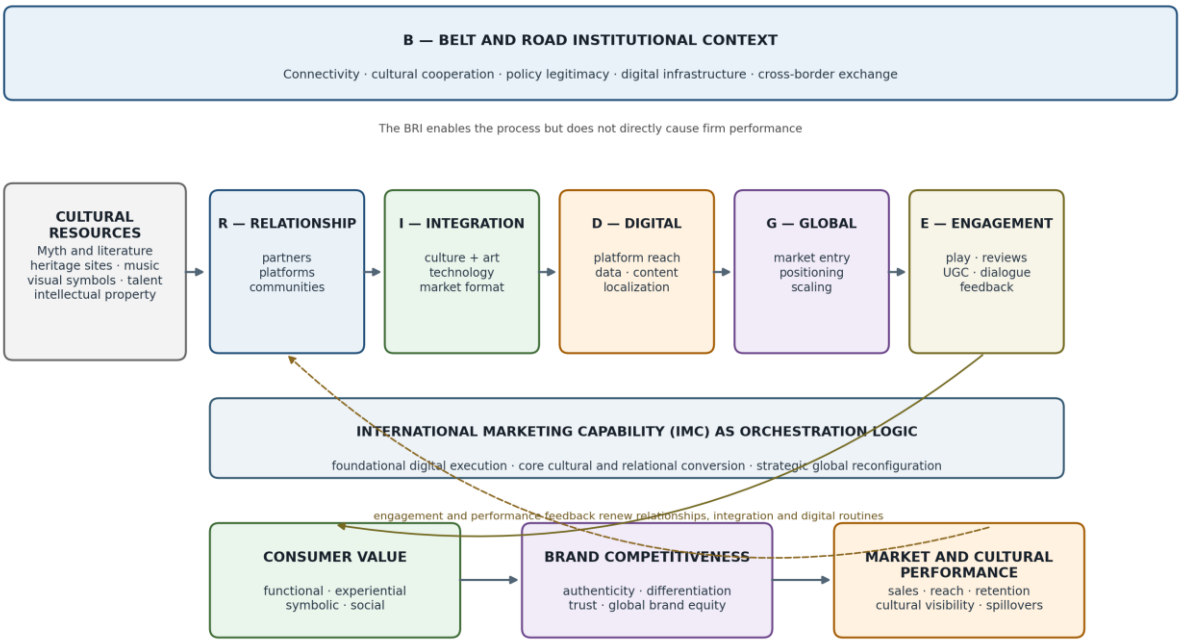


Figure 1 – The BRIDGE–IMC cultural-resource conversion model

Note – developed by the author from the case analysis and international marketing capability literature.

The process is configurational. Strong technology cannot permanently compensate for weak cultural integration; authentic content cannot create international performance without access; global availability cannot sustain a brand without engagement. A bottleneck in one dimension limits the value of the others. The model also distinguishes consumer value from brand competitiveness. Functional, experiential, symbolic and social value strengthen authenticity, differentiation, trust and global brand equity; these brand assets then support sales, retention, portfolio expansion and cultural visibility.

Proposition 7. The BRIDGE dimensions have a complementary, bottleneck-sensitive effect on international marketing capability: weakness in any one of relationship, integration, digital, global or engagement capability constrains the conversion of cultural resources into consumer value.

Proposition 8. Consumer value positively affects international market performance through brand competitiveness, expressed as perceived authenticity, differentiation, trust and global brand equity.

Theoretical contributions

The first contribution is a process extension of international marketing capability theory. Prior research identifies market knowledge, implementation, digital and dynamic capabilities, but it often treats them as a list of firm attributes. BRIDGE–IMC specifies an ordered conversion process and connects macro context, ecosystem relationships, cultural integration, digital execution, strategic expansion and engagement feedback. It thereby shows how ordinary and dynamic capabilities interact in a cultural product.

The second contribution is to place consumer value between cultural resources and performance. Cultural resources are not assumed to possess automatic market value. They become valuable when consumers experience, interpret and socially circulate them. Engagement is both a mediator and a feedback mechanism. This explains why the same product

can achieve technical admiration, cultural pride, symbolic attachment or narrative confusion in different audience groups.

Further conceptual clarification concerns globality. The case separates global availability, global adoption and global diversification. This distinction prevents platform coverage from becoming an inflated measure of international performance and creates more precise indicators for later empirical research.

Managerial and policy implications

Cultural enterprises can use BRIDGE–IMC as a stage-gate diagnostic. Before international launch, managers should identify the cultural resources that form the non-negotiable source identity; map technology, platform, creative and institutional partners; and test whether the integrated product communicates a coherent value proposition. Digital launch plans should connect awareness metrics with evidence of understanding, not only impressions or downloads.

Market dashboards should distinguish access, adoption and diversification. Useful indicators include non-home market revenue, review-language distribution, regional retention, creator participation, cultural comprehension, perceived authenticity and the reuse of jointly created intellectual property. Engagement data should be coded for cultural interpretation and fed back into localization, product design and partner management.

For China–Kazakhstan projects, the model favors co-creation over one-way cultural export. Kazakhstani cultural resources—epics, music, landscapes, nomadic heritage, design and contemporary urban creativity—can be combined with Chinese production, platform and commercialization capabilities, but local creators and institutions should participate in meaning, rights and revenue decisions. Bilateral teams need multilingual boundary spanners and shared rules for IP, data, attribution and audience research.

Policy programs should support the full conversion chain. Events and connectivity correspond mainly to B and R. Sustainable market outcomes also require integration laboratories, localization support, rights services, platform analytics, co-production finance and mechanisms that retain learning after a project ends. Universities and city-based creative hubs in Almaty, Astana and Chinese partner cities can serve as neutral intermediaries.

The Black Myth case also warns against mechanical replication. Its scale, genre and domestic consumer base cannot be copied by most cultural firms. What can be transferred is the capability logic: grounded cultural specificity, complementary relationships, high-quality integration, digital execution, disciplined global expansion and continuous engagement learning.

Conclusion

This article examined how cultural resources can be transformed into global consumer value through international marketing capability. The theory-elaborating case of Black Myth: Wukong produced the BRIDGE–IMC model: Belt and Road context, Relationship, Integration, Digital, Global and Engagement, coordinated by international marketing capability. The case demonstrates that commercial scale followed not from heritage alone, but from the integration of Chinese mythology and material culture with technology, global game conventions, platform access and participatory consumption.

The model’s central mechanism is conversion. Relationships make complementary resources available; integration creates an authentic and accessible proposition; digital capability scales access and intelligence; global capability develops market depth; engagement

produces consumer value and renewal feedback. Brand competitiveness then connects consumer value to market and cultural performance.

The study has limitations. It relies on public archival evidence and cannot observe confidential budgets, regional sales, contracts or internal decisions. Platform metrics change over time, and a single high-performing case cannot represent all cultural industries. The application of the BRI is contextual rather than causal. Future research should conduct interviews with developers, platforms, cultural institutions and players; compare Chinese and Kazakhstani projects; and test the eight propositions through a multi-level design. Consumer surveys can measure functional, experiential, symbolic and social value, while firm-level research can examine capability configurations and non-home market performance.

The case provides a practical conclusion: cultural heritage becomes international market value only when organizations can integrate it, communicate it, invite participation and learn from the meanings consumers return.

References:

1. Auken, B., Kurmashev, A., Ospanova, A., & Kushkumbayev, A. (2025). Creative economy in Central Asia. International Area Studies Review, 28(4), 383–404. https://doi.org/10.69473/iasr.2025.28.4.383

2. Day, G. S. (1994). The capabilities of market-driven organizations. Journal of Marketing, 58(4), 37–52. https://doi.org/10.1177/002224299405800404

3. Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A. S., Kumar, V., Rahman, M. M., Raman, R., Rauschnabel, P. A., Rowley, J., Salo, J., Tran, G. A., & Wang, Y. (2021). Setting the future of digital and social media marketing research: Perspectives and research propositions. International Journal of Information Management, 59, Article 102168. https://doi.org/10.1016/j.ijinfomgt.2020.102168

4. Dyer, J. H., & Singh, H. (1998). The relational view: Cooperative strategy and sources of interorganizational competitive advantage. Academy of Management Review, 23(4), 660–679. https://doi.org/10.5465/amr.1998.1255632

5. Eisenhardt, K. M. (1989). Building theories from case study research. Academy of Management Review, 14(4), 532–550. https://doi.org/10.5465/amr.1989.4308385

6. Elkhoudary, A., Yadav, J. K., & Alam, M. M. D. (2026). Digital marketing capabilities as drivers of SME innovation and performance: A systematic literature review and research agenda for emerging economies. Humanities and Social Sciences Communications, 13, Article 605. https://doi.org/10.1057/s41599-026-07126-x

7. Game Science. (2024, August 23). Black Myth: Wukong has sold 10 million copies across all platforms [Post]. X. https://x.com/BlackMythGame/status/1826985302592049599

8. Gao, Y., Li, J., & He, J. (2025). Hybridizing tradition and technology: Global reception of Black Myth: Wukong as a vehicle for Chinese cultural identity and international media flows. Global Media and China. Advance online publication. https://doi.org/10.1177/20594364251385203

9. Government of the Republic of Kazakhstan. (2025, March 18). Growth in creative industries and the number of employees in this sphere by 15% in Kazakhstan. Official Information Source of the Prime Minister of the Republic of Kazakhstan. https://primeminister.kz/en/news/growth-in-creative-industries-and-the-number-of-employees-in-this-sphere-by-15-in-kazakhstan-29819

10. Guinness World Records. (n.d.). Fastest-selling videogame based on a classic novel. Retrieved July 30, 2026, from https://www.guinnessworldrecords.com/world-records/775494-fastest-selling-videogame-based-on-a-classic-novel

11. Hazzam, J., & Wilkins, S. (2022). International marketing capabilities development: The role of firm cultural intelligence and social media technologies. Journal of Marketing Theory and Practice, 30(3), 325–341. https://doi.org/10.1080/10696679.2021.1946409

12. Hollebeek, L. D., & Macky, K. (2019). Digital content marketing's role in fostering consumer engagement, trust, and value: Framework, fundamental propositions, and implications. Journal of Interactive Marketing, 45, 27–41. https://doi.org/10.1016/j.intmar.2018.07.003

13. Homburg, C., & Wielgos, D. M. (2022). The value relevance of digital marketing capabilities to firm performance. Journal of the Academy of Marketing Science, 50(4), 666–688. https://doi.org/10.1007/s11747-022-00858-7

14. Hoque, M. T., Nath, P., Ahammad, M. F., Tzokas, N., & Yip, N. (2022). Constituents of dynamic marketing capability: Strategic fit and heterogeneity in export performance. Journal of Business Research, 144, 1007–1023. https://doi.org/10.1016/j.jbusres.2022.02.011

15. Kim, S., & Comunian, R. (2024). Higher education and sustainable creative cities in Kazakhstan. Industry and Higher Education, 38(1), 51–63. https://doi.org/10.1177/09504222231222258

16. Klimas, P., Radomska, J., Kościewicz, D., Strzelec, G., & Wrona, S. (2026). What are the drivers, barriers, mechanisms, and anchors of cross-industry cooperation? The perspective of the video game industry ecosystem. Games and Culture, 21(6). https://doi.org/10.1177/15554120251317276

17. Lee, J. T., Lo, W. Y. W., & Abdrasheva, D. (2021). Institutional logic meets global imagining: Kazakhstan's engagement with China's Belt and Road Initiative. Higher Education, 82, 237–253. https://doi.org/10.1007/s10734-020-00634-y

18. Li, F., Larimo, J., & Leonidou, L. C. (2021). Social media marketing strategy: Definition, conceptualization, taxonomy, validation, and future agenda. Journal of the Academy of Marketing Science, 49(1), 51–70. https://doi.org/10.1007/s11747-020-00733-3

19. Li, J., Liu, B., & Qian, G. (2019). The Belt and Road Initiative, cultural friction and ethnicity: Their effects on the export performance of SMEs in China. Journal of World Business, 54(4), 350–359. https://doi.org/10.1016/j.jwb.2019.04.004

20. Li, Q., & Li, X. (2023). Debating the 'Chineseness' of a mobile game in online communities. Global Media and China, 8(4), 442–461. https://doi.org/10.1177/20594364231190313

21. Lim, W. M., & Rasul, T. (2022). Customer engagement and social media: Revisiting the past to inform the future. Journal of Business Research, 148, 325–342. https://doi.org/10.1016/j.jbusres.2022.04.068

22. Mandler, T., Sezen, B., Chen, J., & Özsomer, A. (2021). Performance consequences of marketing standardization/adaptation: A systematic literature review and future research agenda. Journal of Business Research, 125, 416–435. https://doi.org/10.1016/j.jbusres.2020.12.023

23. Manis, K. T., & Madhavaram, S. (2023). AI-enabled marketing capabilities and the hierarchy of capabilities: Conceptualization, proposition development, and research avenues. Journal of Business Research, 157, Article 113485. https://doi.org/10.1016/j.jbusres.2022.113485

24. Mao, H. (2024). A study on cross-cultural communication factors of Black Myth Wukong. In Proceedings of the 2024 3rd International Conference on Social Sciences and Humanities and Arts (pp. 700–708). Atlantis Press. https://doi.org/10.2991/978-2-38476-259-0_73

25. Mattison Thompson, F., & Brouthers, K. D. (2021). Digital consumer engagement: National cultural differences and cultural tightness. Journal of International Marketing, 29(4), 22–44. https://doi.org/10.1177/1069031X211005729

26. Morgan, N. A. (2012). Marketing and business performance. Journal of the Academy of Marketing Science, 40(1), 102–119. https://doi.org/10.1007/s11747-011-0279-9

27. Morgan, N. A., Feng, H., & Whitler, K. A. (2018). Marketing capabilities in international marketing. Journal of International Marketing, 26(1), 61–95. https://doi.org/10.1509/jim.17.0056

28. Morhart, F., Malär, L., Guèvremont, A., Girardin, F., & Grohmann, B. (2015). Brand authenticity: An integrative framework and measurement scale. Journal of Consumer Psychology, 25(2), 200–218. https://doi.org/10.1016/j.jcps.2014.11.006

29. Nieborg, D. B., & Poell, T. (2018). The platformization of cultural production: Theorizing the contingent cultural commodity. New Media & Society, 20(11), 4275–4292. https://doi.org/10.1177/1461444818769694

30. NVIDIA. (2024, March 14). DLSS 3.5 and full ray tracing coming to Black Myth: Wukong, NARAKA: BLADEPOINT and Portal with RTX. https://www.nvidia.com/en-us/geforce/news/gdc-2024-dlss-rtx-full-ray-tracing-game-announcements/

31. Pfajfar, G., Mitręga, M., & Shoham, A. (2024a). International dynamic marketing capabilities: Developments and a research agenda. International Marketing Review, 41(1), 1–41. https://doi.org/10.1108/IMR-10-2023-0283

32. Pfajfar, G., Mitręga, M., & Shoham, A. (2024b). Systematic review of international marketing capabilities in dynamic capabilities view—calibrating research on international dynamic marketing capabilities. International Marketing Review, 41(1), 237–272. https://doi.org/10.1108/IMR-12-2022-0276

33. Rodrigues, C., Brandão, A., Billore, S., & Oda, T. (2024). The mediating role of perceived brand authenticity between brand experience and brand love: A cross-cultural perspective. Journal of Brand Management, 31, 293–309. https://doi.org/10.1057/s41262-023-00342-9

34. State Council Information Office of the People's Republic of China. (2024, October 4). Chinese revel in culture, tourism feast during week-long holiday. https://english.scio.gov.cn/chinavoices/2024-10/04/content_117465788.html

35. Steam. (n.d.). Black Myth: Wukong [Store page]. Retrieved July 30, 2026, from https://store.steampowered.com/app/2358720/Black_Myth_Wukong/

36. Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. Strategic Management Journal, 28(13), 1319–1350. https://doi.org/10.1002/smj.640

37. The Game Awards. (2024). 2024 rewind: Winners. https://thegameawards.com/rewind/year-2024

38. Throsby, D. (2008). The concentric circles model of the cultural industries. Cultural Trends, 17(3), 147–164. https://doi.org/10.1080/09548960802361951

39. Wielgos, D. M., Homburg, C., & Kuehnl, C. (2021). Digital business capability: Its impact on firm and customer performance. Journal of the Academy of Marketing Science, 49(4), 762–789. https://doi.org/10.1007/s11747-021-00771-5

40. Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). SAGE.

41. Yodovich, N., Bagnall, G., Crawford, G., Gislam, C., & Stukoff, M. (2025). Video games in/as culture: The evolving cultural significance of video games. International Journal of the Sociology of Leisure, 8, 339–355. https://doi.org/10.1007/s41978-025-00184-6

UDC 343.2

Ibraimkhan Alikhan
Grade 11 student
Supervisor: Bilimkhan Nazerke
chemistry teacher
Bilim Innovation Lyceum for Gifted Boys
(Karagandy, Kazakhstan)

WHICH CRIMES CAN A 14-YEAR-OLD BE TRIED FOR? THE LIST IN ARTICLE 15 OF KAZAKHSTAN'S CRIMINAL CODE COMPARED WITH 30 OTHER JURISDICTIONS

Abstract: In Kazakhstan criminal responsibility starts at the age of 16, but Article 15 of the Criminal Code lists 31 offences for which a child can be tried from the age of 14. International comparisons usually look only at the minimum age itself, and much less attention is paid to what such lists contain. In this study I compared the Kazakhstan list with the rules of 30 other jurisdictions in Europe and Asia. I read each rule in the current text of the national criminal code, put every listed offence into one of five groups, and traced how the Kazakhstan list changed after the 1997 Code. Eleven of the 31 jurisdictions use a list of offences for younger children, nineteen have one minimum age without exceptions, and Sweden, since September 2026, uses a penalty threshold instead of a list. Among the eleven lists, Kazakhstan's is the second longest (31 offences, median 20) and has the second lowest share of offences against the person (19.4%, median 28.6%), in both cases after Russia. It is also the only list that includes offences of public expression, namely incitement of discord and propaganda of terrorism. When the 1997 list (21 offences) was replaced by the 2014 list (30 offences), all nine new offences were crimes against the state or terrorism offences, while the number of offences against the person stayed at six. Unlike China and Ireland, where a prosecutor must approve the case, and Poland, where a court decides in each case, Kazakhstan applies its list automatically. These results suggest that the list grew mainly together with the offences against the state and terrorism in the Code, and I found no sign that knowledge about how children develop played a role in it.

Keywords: minimum age of criminal responsibility; juvenile justice; Criminal Code of Kazakhstan; comparative law; adolescent development; children's rights

1. Introduction

In Kazakhstan a person can be prosecuted for a crime from the age of 16. Article 15 of the Criminal Code makes an exception: children aged 14 and 15 are criminally responsible for 31 offences named in its second part [1]. Some of these offences are the ones most people would expect, such as murder, rape and kidnapping. Others are less obvious, for example theft of livestock, aggravated hooliganism, damaging items of special value, a knowingly false report of an act of terrorism, and propaganda of terrorism. The list therefore decides which acts of a 14-year-old are treated as crimes and which are not, and it is not obvious from the text why these particular offences were chosen.

The international standard on this question comes from the United Nations Committee on the Rights of the Child. In its General Comment No. 24 (2019) the Committee encourages states to set the minimum age of criminal responsibility at no lower than 14 years, and it bases this on evidence from child development and neuroscience [2]. In the same document the Committee

says it is concerned about rules that allow a lower age when a child is accused of a serious offence, writes that such rules are "usually created to respond to public pressure", and recommends that states abolish them [2]. Kazakhstan's general age of 16 is above the level the Committee asks for, but the list in Article 15 is exactly the kind of exception it describes.

Research on adolescent development explains why the age question is not simple. Adolescents can reason well about hypothetical problems, but their judgement in real situations is less mature than that of adults because of lower responsibility, perspective and self-control [3]. According to the social neuroscience model, the brain systems linked to reward and emotion develop earlier than the systems of cognitive control, which makes risk-taking typical for this age [4], and adolescents take more risks when their peers are present [5]. A study of more than 5,000 people in eleven countries found that cognitive capacity reaches adult levels around the age of 16, while psychosocial maturity continues to develop after 18 [6]. In other words, a 14-year-old may know that an act is forbidden and still be less able than an adult to stop himself, especially in a group.

Most earlier comparative work looks at the general minimum age. Cipriani compared minimum ages around the world from the point of view of children's rights [7], Farmer and Brown and Charles discussed whether the age of 10 in England and Wales can be justified [8, 9], and Llamas and Marinaro used neuroscience to argue for more limited punishment of young offenders in Latin America [10]. Public opinion matters as well: in the Czech Republic, support for a lower minimum age was driven mainly by how people perceive trends in children's crime [11]. When I first planned this study, my question was whether the minimum ages in different countries agree with neuroscience. While reading, I found that this question has already been discussed a lot, so I changed the focus. I could not find a study that compares what the lists of exceptions actually contain, and this is the question of the present paper.

I asked four questions:

• Which of 31 jurisdictions allow a lower age for particular offences, and in what form?

• How long are the lists of offences and what kinds of offences do they contain, and where is Kazakhstan among them?

• How has the Kazakhstan list changed since the 1997 Criminal Code?

• What safeguards come together with these lists?

2. Data and methods

Jurisdictions. I started from a table of 31 jurisdictions that I made from the country pages of the Child Rights International Network (CRIN) [12]. I chose them so that the table would have a wide range: Kazakhstan and its neighbours (Russia, Uzbekistan, Kyrgyzstan, Tajikistan, Turkmenistan, China and Mongolia), seventeen European jurisdictions including those with the lowest and highest ages in Europe (Switzerland, England and Wales, Scotland, Ireland, the Netherlands, Germany, France, Italy, Spain, Poland, Sweden, Norway, Denmark, Belgium, Portugal, Croatia and Greece), and six other countries (Ukraine, Georgia, Turkey, Japan, South Korea and India). The sample is not random. It was chosen to compare Kazakhstan with its region and with countries that are often used as examples in discussions of youth justice.

Checking the rules against the law. CRIN is a useful starting point, but it is a secondary source and some of its pages are several years old. When my first table was checked, it had mistakes. For example, it gave Kazakhstan's ages as "13-14", although Article 15 has no age of 13, and for Uzbekistan it referred to the Russian Criminal Code. Because of this, every rule in the final dataset was read in the current text of the national law, mostly in official legal databases, on 2 October 2026. This mattered: several rules had changed after the CRIN pages

were written. Russia added sabotage offences to its list in 2025 [13], Turkmenistan adopted a new edition of its Criminal Code in force from 1 January 2023 [14], Mongolia amended its list in 2021 [15], Poland added a lower age for aggravated murder by a law of 2022 [16], and Sweden lowered its age for serious offences in September 2026 [17]. For 28 of the 31 jurisdictions the rule was read in the law itself. For Belgium, Greece and Portugal I used secondary legal sources, and none of these three has an offence list.

What I recorded. For every jurisdiction I recorded the general minimum age, whether a lower age applies to particular offences, the lower age, and every offence on the list. An "offence" is one item as the law names it. If the law names several articles together as one item, for example the Ukrainian list names attacks on the life of a state figure, a judge or a police officer together, I counted it once. The law of Uzbekistan lists only article numbers, so there each article is one item.

Five groups of offences. I put every listed offence into one of five groups according to what it protects:

• Against the person: murder, bodily harm, sexual violence, kidnapping, human trafficking and domestic violence.

• Property: theft (including theft of livestock), open theft, robbery, extortion, taking a vehicle, and destroying or damaging property.

• Terrorism, hostage-taking and armed groups: terrorist acts and their financing, propaganda, organisation and training, failure to report them, hostage-taking, false reports of terrorism, attacks on buildings and means of communication, illegal armed groups and banditry.

• The state and international order: attempts on the life of the head of state, officials, judges, police officers or internationally protected persons, sabotage, and incitement of national, religious or other discord.

• Other public safety and order: hooliganism, vandalism, weapons, explosives, drugs, transport, riots, escape from custody, desecration of graves, computer offences, cruelty to animals, arson and poisoning.

Some offences could belong to two groups. I placed incitement of discord (Article 174 of the Kazakhstan Code) in the group of the state and international order, because the Code itself puts it in the chapter on crimes against peace and the security of mankind, next to the attack on internationally protected persons (Article 173) [1]. Hostage-taking was placed with terrorism. To see whether these two choices change the results, I repeated the analysis twice: once with incitement of discord moved to "other", and once with hostage-taking moved to offences against the person.

Offences of public expression. Separately, I marked offences that are committed mainly by saying, writing or publishing something addressed to other people: incitement of hatred or discord, public calls to commit crimes, and propaganda.

History of the Kazakhstan list. I read Article 15 of the 1997 Criminal Code in the text that was in force when the Code was replaced [18], the original text of Article 15 of the 2014 Code [19], and all six laws that amended Article 15 between 2017 and 2026 [1, 20, 21].

Safeguards. For each list I read the age article and the rules it refers to, and recorded whether there is a step that filters cases before prosecution, whether a court decides if the lower age should apply, and what happens to a child with a delay in mental development.

The dataset covers all rules in the chosen jurisdictions, not a sample of cases, so I report counts, percentages and medians and do not use significance tests.

3. Results

3.1 Three ways of setting the age

Table 1 shows all 31 jurisdictions. Nineteen of them use one minimum age without any list. Eleven use a list of offences for which a younger child can be tried, and Sweden uses a third design: since 10 September 2026, a 14-year-old can be punished for offences that carry a minimum penalty of at least four years of imprisonment, as a temporary rule for five years [17]. The Swedish rule does not name the offences; it depends only on how serious the punishment is.

Table 1. Minimum ages of criminal responsibility in 31 jurisdictions

Jurisdiction	General age	Lower age	Offences on the list
Russia	16	14	37
Kazakhstan	16	14	31
Uzbekistan	16	14	22
Tajikistan	16	14	22
Kyrgyzstan	16	14	21
Ukraine	16	14	20
Turkmenistan	16	14	14
Mongolia	16	14	12
Poland	17	15 (14)	10
China	16	14 (12)	8
Ireland	12	10	5
Sweden	15	14	no list, penalty of at least 4 years
India	7	none	none
England and Wales	10	none	none
Switzerland	10	none	none
Netherlands	12	none	none
Scotland	12	none	none
Turkey	12	none	none
France	13	none	none
Croatia	14	none	none
Georgia	14	none	none
Germany	14	none	none
Italy	14	none	none
Japan	14	none	none
South Korea	14	none	none
Spain	14	none	none
Denmark	15	none	none
Greece	15	none	none
Norway	15	none	none
Portugal	16	none	none
Belgium	18	none	none

The lower age applies only to the listed offences. China also allows prosecution from 12 for two offences with the approval of the Supreme People's Procuratorate, and Poland from 14 for aggravated murder. In the Netherlands, Belgium and some other countries older children can be treated as adults for serious offences, which is a different rule and was not counted as a list. In France, Turkey and India children are responsible above the general age shown here only if they are found able to understand their acts (until 13, 15 and 12 years respectively). Sources: national criminal codes, read on 2 October 2026; for Belgium, Greece and Portugal, secondary sources.

The lists are concentrated in one region. All eight post-Soviet jurisdictions in the sample except Georgia have a list, and so do China and Mongolia. Georgia also had a list for a short time: amendments adopted in 2007 lowered the age to 12 for premeditated murder, intentional damage to health, rape, most types of robbery and possession of knives from 1 July 2008, and after criticism from the Committee on the Rights of the Child the single age of 14 was restored in 2010 [22, 23]. In Europe only Poland and Ireland use lists. I also checked two post-Soviet countries outside my sample, Belarus and Armenia, and both of them use the same kind of list with the age of 14 [24, 25].

3.2 How long the lists are and what is on them

Figure 1 shows the eleven lists divided into the five groups. They range from 5 offences in Ireland to 37 in Russia, and the median is 20. Kazakhstan has the second longest list with 31 offences.

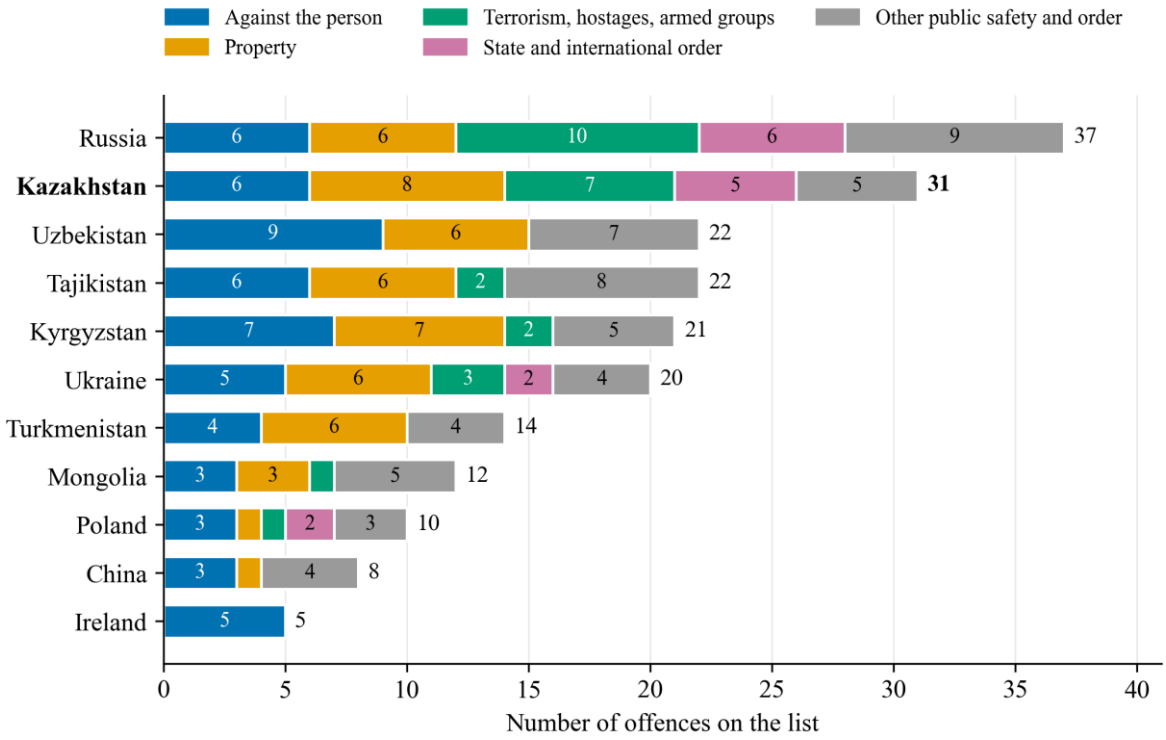


Figure 1. Offences for which a child below the general age can be tried, by group, in the eleven jurisdictions that use a list

The number of offences against the person is similar in most lists: six in Kazakhstan, Russia and Tajikistan, seven in Kyrgyzstan and nine in Uzbekistan. What makes the Kazakhstan and Russia lists longer is everything else. Offences against the person are 19.4% of the Kazakhstan list and 16.2% of the Russian list. These are the two lowest shares among the eleven lists, while the median is 28.6%. In Ireland, the list includes only offences against the person.

The two long lists also contain most of the terrorism and state offences in the whole dataset. Russia has 10 terrorism offences and 6 state offences, and Kazakhstan has 7 and 5. Ukraine has 3 and 2, Poland 1 and 2, and the other seven lists have at most 2 terrorism offences and no state offences at all. In Kazakhstan these two groups together make up 12 of the 31 offences (38.7%), which is twice as many as the six offences against the person.

The two checks with different coding did not change this picture. With incitement of discord moved to "other", Kazakhstan has 4 state offences instead of 5, and its share of offences against the person does not change. With hostage-taking counted as an offence against the

person, Kazakhstan's share rises to 22.6% and Russia's to 18.9%, and they are still the two lowest.

3.3 Offences of public expression

The Kazakhstan list is the only one of the eleven that includes offences of public expression. A 14-year-old in Kazakhstan can be tried for incitement of social, national, ethnic, racial or religious discord (Article 174) and for propaganda of terrorism or public calls to commit an act of terrorism (Article 256) [1]. Russia has a similar offence of public calls to terrorist activity and propaganda of terrorism (Article 205.2 of its Code), but it is not on the Russian list, so in Russia this offence applies only from 16 [13]. Kazakhstan and Russia both include a knowingly false report of an act of terrorism, which is also an offence committed with words, but it is about a specific false message rather than about views.

3.4 How the Kazakhstan list changed

Figure 2 compares three versions of the Kazakhstan list. The 1997 Code had 21 offences in it when it was replaced, and six of them were offences against the person [18]. The 2014 Code started with 30 offences [19]. All nine new ones were crimes against the state or terrorism offences: the attack on internationally protected persons, incitement of discord, the attempts on the life of the First President and of the President, sabotage, propaganda of terrorism, creating or taking part in a terrorist group, financing terrorism or extremism, and attacks on buildings and means of communication. No new offence against the person was added, so their share fell from 28.6% to 20.0%. It is interesting that 28.6% is exactly the median of the eleven lists today.

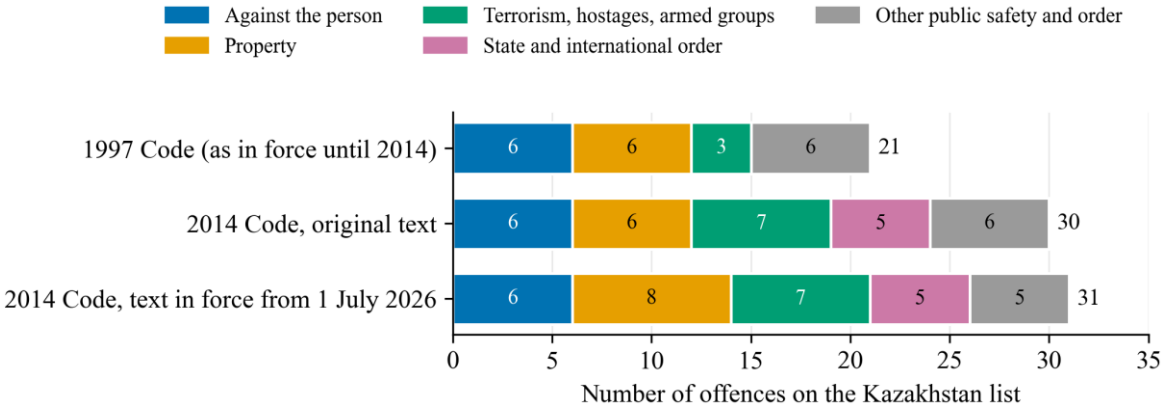


Figure 2. The Kazakhstan list in the 1997 Criminal Code, in the original text of the 2014 Criminal Code and in the text in force from 1 July 2026

After 2014 the list changed little. In 2019 livestock theft was added and vandalism was removed [20], and in 2024 the destruction or damage of items of special value was added [21]. The amendments of 2017, 2020, 2025 and 2026 changed only the wording, for example "the First President of the Republic of Kazakhstan, Elbasy" became "the ex-President of the Republic of Kazakhstan" in 2025 [1].

One change went in the other direction. The 1997 Code freed a minor with a delay in mental development from responsibility only for crimes of small or medium gravity [18]. In the 2014 Code this rule applies to any offence [19].

3.5 Safeguards

Table 2 shows what each list comes with. In three jurisdictions there is a step before the case or a decision of a court. In Ireland, when a child under 14 is charged, no further proceedings can take place without the consent of the Director of Public Prosecutions [26]. In China, a child aged 12 or 13 can be prosecuted only for intentional homicide or intentional injury that causes

death or, by especially cruel means, serious disability, and only with the approval of the Supreme People's Procuratorate [27]. In Poland, the court decides whether a child aged 14 to 16 should be tried under the adult rules, taking into account the circumstances of the case and the child's development and personal conditions [16].

Table 2. Safeguards that come with the lists

Jurisdiction	Safeguard linked to the lower age	Type
Ireland	Proceedings against a child under 14 need the consent of the Director of Public Prosecutions	Filter before prosecution
China	Prosecution at 12-13 needs the approval of the Supreme People's Procuratorate	Filter before prosecution
Poland	The court decides whether the lower age is applied, considering the child's development	Decision of the court
Kazakhstan, Russia, Tajikistan	No responsibility if a delay in mental development, not caused by a mental disorder, prevented the child from fully understanding the act	Defence during the case
Mongolia	Punishment may be waived if a delay in mental development prevented full understanding	At sentencing
Turkmenistan	A delay in mental development only reduces the punishment	At sentencing
Ukraine, Uzbekistan, Kyrgyzstan	Nothing in the article on age	None

Based on the article on the age of criminal responsibility and the rules it refers to in each code.

Kazakhstan belongs to the group with a defence for delay in mental development. This rule protects some children, but it works inside a case that has already started. It does not decide whether a 14-year-old should be prosecuted in the first place, which is what the Irish and Chinese filters do.

4. Discussion

The main result of this study is that Kazakhstan's list is long, and that it is long because of offences that are not about violence against people. Kazakhstan has the same number of offences against the person on its list as Russia and Tajikistan, and fewer than Uzbekistan or Kyrgyzstan. The difference comes from terrorism, state and property offences. The comparison with the 1997 Code shows where this difference came from: all nine new offences in 2014 were from these groups.

I cannot say from the text of the law why each offence was added, because I did not find the explanations that were written for the 2014 Code. Still, the pattern fits what the Committee on the Rights of the Child writes about such lists, namely that they usually respond to public pressure and are not based on an understanding of how children develop [2]. The Czech study of public opinion shows that support for a lower minimum age depends on how people see trends in children's crime [11]. Russia did something similar very recently: in November 2025 it added sabotage and three related offences to its list, and this was done by a law about sabotage [13]. In Kazakhstan the nine new offences came with the new Code in 2014, at the same time as the Code's chapters on terrorism and crimes against the state.

The developmental research is relevant here. Several offences on the Kazakhstan list can be committed in a group or online, such as aggravated hooliganism, incitement of discord and

spreading propaganda. Teenagers take more risks when their peers are around [4, 5], and their psychosocial maturity is still developing at 14 and 15 even if their reasoning is close to adult level [6]. A 14-year-old who shares extremist material in a group chat may understand that it is wrong and still do it to look brave in front of friends. The act can still be harmful, but the reasons the Committee gives for a minimum age of at least 14 seem to apply especially strongly to offences like these.

The comparison also shows that the direction of change is not the same everywhere. Georgia lowered its age for a list of offences in 2008 and returned to a single age in 2010 [22, 23], and Kazakhstan itself made its delay-in-development rule wider in 2014. At the same time, Poland added a lower age for aggravated murder by a law of 2022 [16], and Sweden lowered its age for serious offences in 2026 [17]. So the pressure to treat younger children as adults for serious crimes exists in Europe too. The post-Soviet lists, and especially the Kazakhstan list, are much wider. The Polish and Swedish exceptions are limited to the most serious crimes, and in Poland a court decides in each case whether to apply the lower age.

In my opinion, the age of 14 is easiest to defend for offences where a person is killed or seriously hurt, because the harm is very serious and the child can understand it. For offences such as livestock theft or damaging property, and especially for offences of expression, it is harder to see why a 14-year-old should be treated as an adult when the same Code treats him as a child for almost everything else.

Based on this comparison, three changes could be discussed. First, the list could be reviewed with the General Comment in mind, starting with the offences of public expression and the property offences that do not involve violence. Second, for the offences that stay on the list, Kazakhstan could add a filter before prosecution, similar to the consent of the prosecutor in Ireland or the approval of the highest prosecutor's office in China. Third, it would help to publish how often each part of Article 15 is used against 14- and 15-year-olds, because without these numbers it is impossible to know what the list means in practice. These are ideas that follow from the comparison, and I did not test them. In September 2026 the Committee on the Rights of the Child considered the combined fifth and sixth reports of Kazakhstan. Its concluding observations, published on 28 September 2026 as an advance unedited version, do not mention Article 15 or the list. They do ask Kazakhstan to establish diversion mechanisms and other measures outside the courts for children accused of crimes, and to make sure that any restrictions of children's freedom of expression under the laws on extremism, false information or public order are necessary and proportionate [28]. The first point is close to the idea of a filter before prosecution, and the second one concerns the same kind of offences as Articles 174 and 256.

5. Limitations

This study compares rules in the law, not how they are used. A long list does not tell how many 14- and 15-year-olds are actually prosecuted for each offence, and I did not have these statistics. The sample of 31 jurisdictions was chosen by me and is not random, so the medians describe this sample and not the whole world. Putting offences into groups requires judgement. I tested the two most debatable choices and the main results did not change, but another person could still group some offences differently. Lists are counted in different units in different laws, and Uzbekistan's list is written as article numbers, so the lengths are not perfectly comparable. For Belgium, Greece and Portugal I relied on secondary sources. Several laws were read in Russian or English translations or in automatic translation, which may miss details. Finally, laws change often, as the 2025 and 2026 changes in Russia, Kazakhstan and Sweden show, so the dataset describes the situation on 2 October 2026.

6. Conclusion

Eleven of 31 jurisdictions allow a younger child to be tried for a list of offences, and Kazakhstan's list is one of the longest of them, with the second lowest share of offences against the person and the only offences of public expression in the sample. Its growth from 21 to 30 offences in 2014 came entirely from crimes against the state and terrorism offences. Kazakhstan also applies its list automatically, while in China and Ireland a prosecutor must approve the case and in Poland a court decides in each case. If the age of 14 is kept for the most serious violent crimes, the rest of the list could be reviewed, and a filter before prosecution could be added.

Author contribution and use of AI

The author chose the topic, made the first table of 31 jurisdictions from CRIN, and wrote the opinion in the Discussion. The author's mentor checked every rule against the current national laws, collected the history of the Kazakhstan list, built and coded the dataset, made the figures, and helped write and edit the English text. Generative AI tools were used to help with translation and editing. Every legal provision and number in the paper was checked against its source by the author's mentor.

Data availability

The coded dataset with all 202 listed offences and the source of every rule is available from the author on request.

References:

1. Criminal Code of the Republic of Kazakhstan No. 226-V of 3 July 2014 (as amended up to the Law No. 307-VIII of 11 June 2026) [in Russian]. URL: https://adilet.zan.kz/rus/docs/K1400000226 (accessed: 02.10.2026).

2. Committee on the Rights of the Child. General comment No. 24 (2019) on children's rights in the child justice system. CRC/C/GC/24. Geneva: United Nations, 18 September 2019.

3. Cauffman E., Steinberg L. (Im)maturity of judgment in adolescence: why adolescents may be less culpable than adults // Behavioral Sciences & the Law. 2000. Vol. 18, No. 6. P. 741-760. DOI: 10.1002/bsl.416.

4. Steinberg L. A social neuroscience perspective on adolescent risk-taking // Developmental Review. 2008. Vol. 28, No. 1. P. 78-106. DOI: 10.1016/j.dr.2007.08.002.

5. Smith A.R., Chein J., Steinberg L. Peers increase adolescent risk taking even when the probabilities of negative outcomes are known // Developmental Psychology. 2014. Vol. 50, No. 5. P. 1564-1568. DOI: 10.1037/a0035696.

6. Icenogle G., Steinberg L., Duell N., Chein J., Chang L., Chaudhary N., et al. Adolescents' cognitive capacity reaches adult levels prior to their psychosocial maturity: evidence for a "maturity gap" in a multinational, cross-sectional sample // Law and Human Behavior. 2019. Vol. 43, No. 1. P. 69-85. DOI: 10.1037/lhb0000315.

7. Cipriani D. Children's Rights and the Minimum Age of Criminal Responsibility: A Global Perspective. London: Routledge, 2016. DOI: 10.4324/9781315571584.

8. Farmer E. The age of criminal responsibility: developmental science and human rights perspectives // Journal of Children's Services. 2011. Vol. 6, No. 2. P. 86-95. DOI: 10.1108/17466661111149385.

9. Brown A., Charles A. The minimum age of criminal responsibility: the need for a holistic approach // Youth Justice. 2021. Vol. 21, No. 2. P. 153-171. DOI: 10.1177/1473225419893782.

10. Llamas N.E., Marinaro J.A. Neuroscience in youth criminal law: reconsidering the measure of punishment in Latin America // Frontiers in Psychology. 2020. Vol. 11. Art. 302. DOI: 10.3389/fpsyg.2020.00302.

11. Hulmakova J., Tomasek J., Krulichova E. Public opinion on the minimum age of criminal responsibility: the role of cognition and emotions // International Criminal Justice Review. 2025. Vol. 35, No. 3. P. 330-347. DOI: 10.1177/10575677241282285.

12. Child Rights International Network. Minimum ages of criminal responsibility in Asia; Minimum ages of criminal responsibility in Europe. URL: https://archive.crin.org/en/home/ages/asia.html; https://archive.crin.org/en/home/ages/europe.html (accessed: 19.09.2026).

13. Criminal Code of the Russian Federation No. 63-FZ of 13 June 1996, Articles 20 and 205.2 (as amended up to 04.08.2026, including Federal Law No. 420-FZ of 17 November 2025) [in Russian]. URL: https://www.consultant.ru/document/cons_doc_LAW_10699/ (accessed: 02.10.2026).

14. Criminal Code of Turkmenistan (new edition, in force from 1 January 2023), Article 19 [in Russian]. URL: https://mejlis.gov.tm/single-code/27?lang=ru (accessed: 02.10.2026).

15. Criminal Code of Mongolia of 3 December 2015, Article 6.2 (as amended on 17 December 2021) [in Mongolian]. URL: https://legalinfo.mn/mn/detail?lawId=11634 (accessed: 02.10.2026).

16. Penal Code of the Republic of Poland of 6 June 1997, Article 10, consolidated text Dz.U. 2025 poz. 383; Act of 7 July 2022 amending the Penal Code, Dz.U. 2022 poz. 2600 [in Polish]. URL: https://api.sejm.gov.pl/eli/acts/DU/2025/383 (accessed: 02.10.2026).

17. Brottsbalk (1962:700) [Swedish Penal Code], Chapter 1, Section 6, as amended by Lag (2026:1528), temporary text in force until 10 September 2031 [in Swedish]. URL: https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/brottsbalk-1962700_sfs-1962-700/ (accessed: 03.10.2026).

18. Criminal Code of the Republic of Kazakhstan No. 167-I of 16 July 1997 (no longer in force), Article 15 [in Russian]. URL: https://adilet.zan.kz/rus/docs/K970000167_ (accessed: 02.10.2026).

19. Criminal Code of the Republic of Kazakhstan No. 226-V of 3 July 2014, original text of 03.07.2014 [in Russian]. URL: https://old.adilet.zan.kz/rus/docs/K1400000226/03.07.2014 (accessed: 02.10.2026).

20. Law of the Republic of Kazakhstan No. 292-VI of 27 December 2019 on amendments to some legislative acts on improving criminal and criminal procedure legislation and strengthening the protection of individual rights [in Russian]. URL: https://adilet.zan.kz/rus/docs/Z1900000292 (accessed: 02.10.2026).

21. Law of the Republic of Kazakhstan No. 111-VIII of 5 July 2024 on amendments to some legislative acts [in Russian]. URL: https://adilet.zan.kz/rus/docs/Z2400000111 (accessed: 02.10.2026).

22. Human Rights Watch. Georgia: Raise Age of Criminal Responsibility. 8 June 2008. URL: https://www.hrw.org/news/2008/06/08/georgia-raise-age-criminal-responsibility (accessed: 02.10.2026).

23. Criminal Code of Georgia of 22 July 1999, Article 33 (as amended on 23 February 2010) [in Russian translation]. URL: https://matsne.gov.ge/ru/document/view/16426 (accessed: 02.10.2026).

24. Criminal Code of the Republic of Belarus No. 275-Z of 9 July 1999, Article 27 [in Russian]. URL: https://pravo.by/document/?guid=3871&p0=hk9900275 (accessed: 02.10.2026).

25. Criminal Code of the Republic of Armenia No. ZR-199 of 27 May 2021, Article 19 [in Russian translation]. URL: https://old.parliament.am/law_docs_8/27052021HO199_rus.pdf (accessed: 02.10.2026).

26. Children Act 2001 (Ireland), section 52, as amended by the Criminal Justice Act 2006, section 129. Revised Acts, Law Reform Commission. URL: https://revisedacts.lawreform.ie/eli/2001/act/24/section/52/revised/en/html (accessed: 02.10.2026).

27. Amendment XI to the Criminal Law of the People's Republic of China, adopted on 26 December 2020, item 1 (Article 17) [in Chinese]. URL: https://www.moj.gov.cn/pub/sfbgw/jgsz/jgszzsdw/zsdwflyzzx/flyzzxzcxx/zcxxzcfg/zcfgfl/202012/t20201227_188562.html (accessed: 02.10.2026).

28. Committee on the Rights of the Child. Concluding observations on the combined fifth to sixth periodic reports of Kazakhstan. CRC/C/KAZ/CO/5-6 (advance unedited version). 28 September 2026. Paras. 19 and 52. URL: https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=CRC%2FC%2FKAZ%2FCO%2F5-6&Lang=en (accessed: 03.10.2026).

UDC 339.5

Yerboluly Madi
Grade 11 student
Supervisor: Iembergen Assem
School-Lyceum No. 119
(Almaty, Kazakhstan)

DID IMPORTED CARS REALLY GET CHEAPER? CHINESE CARS AND THE AVERAGE CUSTOMS VALUE OF PASSENGER CARS IMPORTED INTO KAZAKHSTAN, 2015-2025

Abstract: Chinese cars have become the largest part of Kazakhstan's car imports, and a common explanation is that they are simply cheap. In this study I used UN Comtrade data on passenger cars (Harmonized System heading 8703) to check how China's share and the average customs value of an imported car changed between 2015 and 2025. According to Kazakhstan's records, China supplied 2.0% of imported passenger cars in 2015 and 69.1% in 2025, and the average value of an imported unit fell by 39% between 2023 and 2025. However, three checks showed that most of this fall does not describe real cars. First, in 2025 Kazakhstan recorded 72,366 units, most of them from China, that weighed about 180 kg and were worth about 670 US dollars each, which is not possible for a passenger car. China itself recorded no exports to Kazakhstan in this subheading, but recorded 65,688 motorcycles of 250-500 cm³, so these units are most likely light vehicles classified differently by the two countries. Without them, China's share of cars was 56.5% and the average value fell by 16%, while the average weight of a car did not change. Second, inside the same engine groups, Chinese cars became about 12% cheaper in Kazakhstan's records and about 17% cheaper in China's own records, so this part of the fall is real. Third, for South Korea and Germany the average value fell by 67% and 47% in Kazakhstan's records but rose by 9% and 11% in the exporters' records. The results show that Kazakhstan's car import statistics should not be used without checking subheadings and mirror data, and that the idea that Chinese cars pushed prices down explains only a small part of the change.

Keywords: car imports; unit values; mirror statistics; UN Comtrade; China; Kazakhstan

1. Introduction

Before I started this project I helped for some time in my uncle's company, which brings cars to Kazakhstan from China, Korea, Germany and other countries, and I filled in customs declarations there. Every declaration has the same fields: where the car was made, its engine, its weight and its customs value. When I later looked at Kazakhstan's car imports as a whole, the most visible change was the rise of Chinese cars. In Kazakhstan's records, China supplied 2.0% of imported passenger cars in 2015 and 69.1% in 2025 [1].

A common explanation is that Chinese cars are cheap. If this were the whole story, the average customs value of a car from China should fall, while cars from other countries keep their values. When my mentor and I first looked at the data, we found something different. Between 2023 and 2025 the average value per imported car fell not only for China but also for South Korea and Germany, and for Korea it fell even more. This made the question more interesting than "Chinese cars are cheap".

Economists often use unit values, the value of trade divided by its quantity, as a rough measure of the price and quality of traded goods [2, 3]. A unit value can fall for different reasons. The goods themselves can become cheaper. The mix can change, for example from large to small cars or from new to used ones. The goods can be counted or classified differently. Or the declared values can be lower than the real ones, which is a known problem in customs statistics and can be checked by comparing what the exporting country reports with what the importing country reports [4, 5]. Kazakhstan's car market also changed a lot in these years. After 2022 part of the trade with Russia went through Central Asia [6], and from 1 December 2024 Kazakhstan changed the rules for importing cars, so that only official dealers can import new cars and other importers can bring only cars older than three years [7].

In this study I asked four questions:

• How did China's share of passenger cars imported into Kazakhstan change between 2015 and 2025?

• How did the average customs value of an imported car change, and was this change specific to China?

• Was the fall caused by a change in what kind of cars were imported (engine type and size) or by lower values within the same kind of car?

• Do the exporting countries report the same numbers and values as Kazakhstan?

2. Data and methods

Data. I used the UN Comtrade database, which collects the official trade statistics that countries report to the United Nations [1]. I took Kazakhstan's imports of passenger cars, which are heading 8703 of the Harmonized System [8], for every year from 2015 to 2025. For each year I downloaded the number of units, the customs value in US dollars and the net weight, both for all countries together and for five main partners: China, South Korea, Japan, Germany and Russia. Import values are CIF values, which include transport and insurance to the border. I downloaded all data with Python through the Comtrade API, one year and one partner at a time.

A technical trap. For one year and one partner, Comtrade gives several rows: a total row and separate rows by mode of transport and customs procedure. If all rows are added together, the result is about twice too large. My first attempt gave 224 cars from China in 2018 instead of 112. To avoid this, I kept only the total rows (mode of transport 0, customs procedure C00 and second partner 0).

Engine groups. For 2023-2025 I also downloaded the six-digit subheadings of heading 8703, which divide cars by engine type and engine volume. I grouped them into petrol cars up to 1.0 litre, 1.0-1.5 litres, 1.5-3.0 litres and over 3.0 litres, diesel cars, hybrids, plug-in hybrids and electric cars. Subheading 8703 10 covers special vehicles such as snowmobiles and golf cars.

Light units. While checking the subheadings, I found a large number of units in subheadings 8703 10 and 8703 21 that were much lighter and cheaper than any car (section 3.2). I call them "light units" and report the results both with and without them. In the main analysis of values I use only the other subheadings and call them "cars". For Korea, Japan and Germany these two subheadings contain very few units, so this choice changes almost nothing for them.

Average value and its change. The average value per car is the total value divided by the number of cars. To separate the effect of a different mix of cars from the effect of lower values within the same type of car, I used a shift-share decomposition. The change in the average value between 2023 and 2025 was split into a mix effect, which comes from changes in the shares of engine groups, and a within effect, which comes from changes in the average value

inside each group. I used the average of the two years as weights, so the two parts add up exactly to the total change.

Mirror data. For each partner I also downloaded what that country reports as its exports of passenger cars to Kazakhstan. Exports and imports should describe the same cars, but there are known reasons for small differences: export values are usually FOB values without transport costs, so import values should be a little higher, and the importing country usually records the country where the goods were made, while the exporting country records the country of destination [5]. Russia has not reported its trade to Comtrade since 2022, so there is no mirror data for Russia in these years.

Re-exports. Finally, I downloaded Kazakhstan's exports of passenger cars to the world and to Russia, Kyrgyzstan and Uzbekistan, to see whether some imported cars left the country again.

3. Results

3.1 China's share and the quality of the data

In Kazakhstan's records, China's share of imported passenger cars was 2.0% in 2015 (1,534 of 76,931 units), 35.1% in 2023, 47.0% in 2024 and 69.1% in 2025 (163,944 of 237,338 units). By value, China's share in 2025 was 49.6%.

These numbers cannot be read as a smooth trend, because Kazakhstan's own data for the middle years are incomplete. In 2015 and 2016 Kazakhstan's numbers are close to what its partners reported, for example 2,298 cars from Korea in Kazakhstan's records and 2,303 in Korea's. From 2017 to 2019, Kazakhstan recorded only about 90-130 cars a year from China and 60-360 from Korea, while China reported 2,720-3,390 and Korea 1,081-4,700 cars sent to Kazakhstan. For 2020-2022 Kazakhstan reported cars from Russia only. Figure 1 shows this by comparing China's own exports to Kazakhstan with Kazakhstan's imports from China. China's exports grew from 4,919 cars in 2021 to 20,446 in 2022 and 207,002 in 2025, so the rise of Chinese cars is real, but its speed in the middle years can be seen only in China's data. For this reason, I compare values only for 2023-2025, when Kazakhstan's data cover all partners.

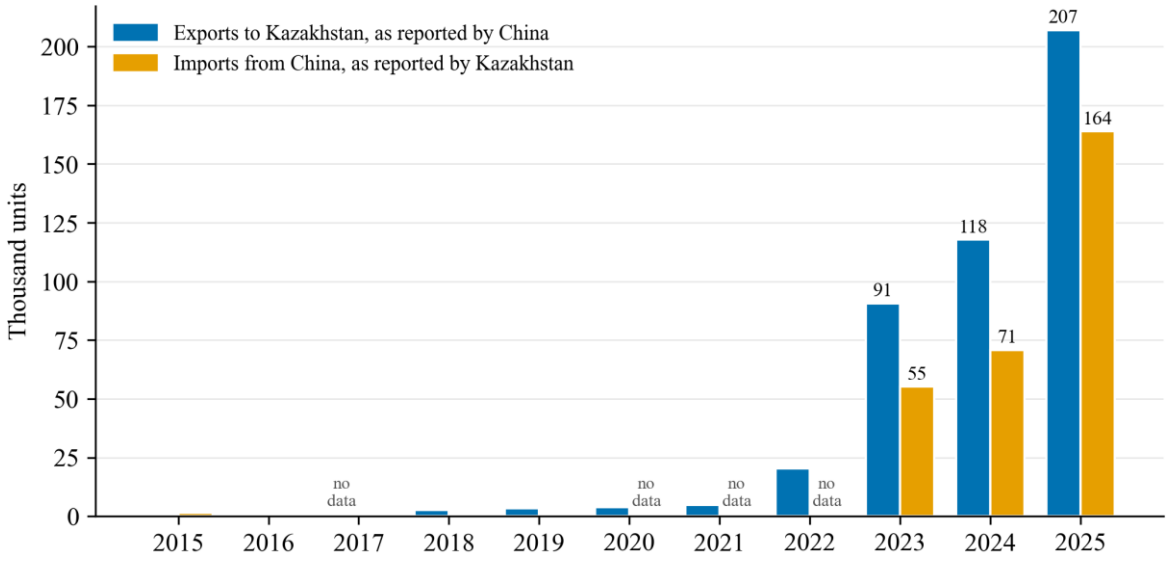


Figure 1. Passenger cars (heading 8703) that China reported as exported to Kazakhstan and that Kazakhstan reported as imported from China, 2015-2025

3.2 Light units that are not cars

Between 2023 and 2025 the average value of one unit imported into Kazakhstan fell by 39%, from 19,477 to 11,947 US dollars. When I looked at the subheadings, most of this fall came from one place. In 2025 Kazakhstan recorded 66,944 units from China in subheading 8703 21 (petrol cars with an engine up to 1.0 litre) with an average weight of 179 kg and an average value of 437 dollars. Together with subheading 8703 10, Kazakhstan recorded 72,366 light units from all countries in 2025, compared with 5,710 in 2023, and 70,754 of them came from China (Figure 2). A passenger car usually weighs more than 800 kg, so these units cannot be ordinary cars.

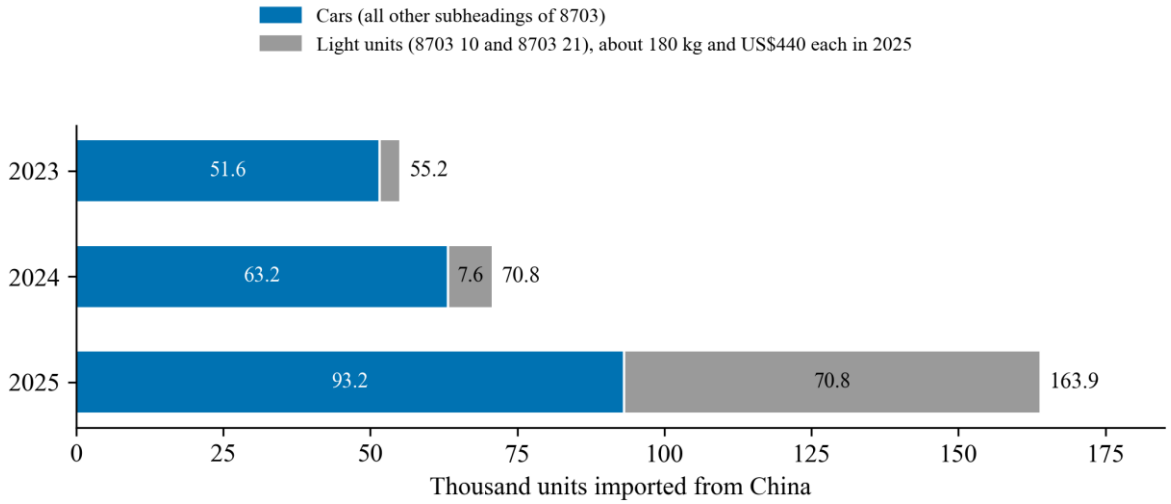


Figure 2. Units imported into Kazakhstan from China in heading 8703: cars and light units, 2023-2025

China's records show where these units probably come from. China reported no exports of passenger cars to Kazakhstan in subheading 8703 21 in any of these years. At the same time, China's exports to Kazakhstan of motorcycles with an engine of 250-500 cm³ (subheading 8711 30) rose from 1,248 units in 2023 to 65,688 in 2025, with an average weight of 127 kg, while Kazakhstan recorded only 6,522 such motorcycles from China in 2025. The two numbers, 65,688 and 66,944, are very close and appeared in the same year. The most likely explanation is that the same light vehicles, for example all-terrain vehicles, are recorded as motorcycles by China and as small passenger cars by Kazakhstan. I cannot prove this from Comtrade data alone, but the weight and the value make it clear that these units should not be counted as cars.

Without the light units, China supplied 34.1% of cars in 2023, 44.7% in 2024 and 56.5% in 2025, and 49.3% of their value in 2025. The average value of a car fell by 16%, from 20,046 to 16,895 dollars, instead of 39%. The average weight of a car stayed almost the same: 1,674 kg in 2023 and 1,669 kg in 2025. Table 1 shows the same comparison for each partner.

Table 1. Cars imported into Kazakhstan by partner, 2023 and 2025, without light units

Partner	Cars, 2023	Cars, 2025	US$ per car, 2023	US$ per car, 2025	kg per car, 2025
China	51,575	93,190	16,750	14,752	1,594
South Korea	19,628	12,396	13,204	4,401	1,636
Japan	27,846	22,284	31,446	35,122	1,938
Germany	4,676	3,722	49,579	26,335	1,888

Russia	416	6,438	13,819	9,605	1,201
All countries	151,286	164,972	20,046	16,895	1,669

Source: UN Comtrade, Kazakhstan's import records, heading 8703 without subheadings 8703 10 and 8703 21. Values are CIF.

3.3 A different mix or lower values?

Table 2 shows the decomposition of the change in the average value of a car between 2023 and 2025. For all countries together, the average value of a car fell by 3,263 dollars. Only 915 dollars of this came from a different mix of engine groups, and 2,348 dollars came from lower values inside the same groups. In other words, most of the fall happened because the same kind of car was recorded as cheaper, not because people started to import smaller cars. This agrees with the stable average weight.

Table 2. Change in the average value of a car between 2023 and 2025, split into the mix effect and the within effect (US$)

Partner	Records of	US$ per car, 2023	US$ per car, 2025	Mix effect	Within effect
China	Kazakhstan	16,750	14,752	+1,717	-3,715
China	China	18,146	15,130	+3,003	-6,019
South Korea	Kazakhstan	13,204	4,401	-1,845	-6,958
South Korea	South Korea	29,739	32,422	-2,016	+4,699
Germany	Kazakhstan	49,579	26,335	-6,158	-17,085
Germany	Germany	82,274	91,477	+2,296	+6,906
Japan	Kazakhstan	31,446	35,122	-3,007	+6,682
Japan	Japan	32,270	36,297	-520	+4,547
All countries	Kazakhstan	19,987	16,724	-915	-2,348

The values are calculated from the subheadings, so for all countries they differ slightly from Table 1, where the total of heading 8703 is used. Exporters' values are FOB.

For China the two parts go in opposite directions. The mix effect was positive: plug-in hybrids and electric cars, which are more expensive, grew from about 2% of Chinese cars in 2023 to 21% in 2025. But inside each group Chinese cars became cheaper. China's own export records show the same pattern, and even a larger fall inside the groups. For example, in China's records the average value of a plug-in hybrid exported to Kazakhstan fell from 51,257 dollars in 2023 to 24,128 dollars in 2025, and that of an electric car from 40,846 to 19,161 dollars. So for Chinese cars the fall in value is real and can be seen from both sides.

3.4 What the exporters say

Figure 3 compares the average value of a car in Kazakhstan's records and in the exporters' records. For China and Japan the two sources are close and move in the same direction. For South Korea and Germany they move in opposite directions. In Kazakhstan's records, the average Korean car fell from 13,204 to 4,401 dollars (-67%) and the average German car from 49,579 to 26,335 dollars (-47%). In Korea's and Germany's own records, the average value of a car exported to Kazakhstan rose by 9% and 11%.

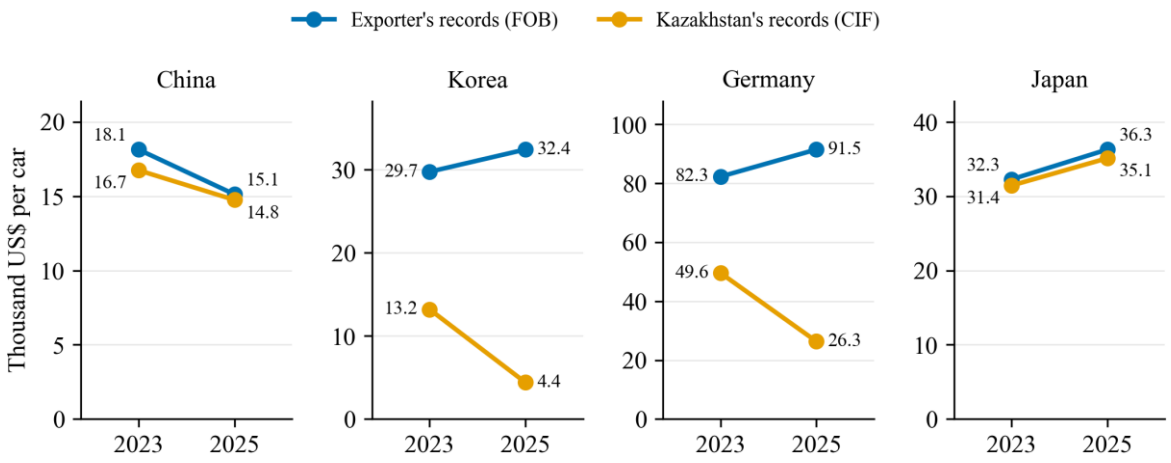


Figure 3. Average value of a car (without light units) imported into Kazakhstan, in the exporter's records and in Kazakhstan's records, 2023 and 2025

The numbers of cars also differ. In 2025 Korea reported 52,070 cars exported to Kazakhstan, but Kazakhstan recorded 12,485 cars from Korea. China reported 207,002, and Kazakhstan 163,944. For Germany and Japan the difference goes the other way: Kazakhstan recorded more cars of German and Japanese origin (3,727 and 22,294) than Germany and Japan reported sending to Kazakhstan (1,859 and 16,375). The Korean cars that Kazakhstan recorded were real cars, with an average weight of 1,636 kg, but in 2025 an average Korean car with a 1.5-3.0 litre engine was declared at only 3,886 dollars, while Korea reported 29,839 dollars for this group.

3.5 Cars that left Kazakhstan again

Kazakhstan's own exports of passenger cars also changed. Exports to Russia rose from 2,073 cars in 2022 to 8,882 cars worth 193 million dollars in 2023, and then fell to 381 cars in 2024 and 163 in 2025. In 2023 these exports were equal to about 6% of the cars Kazakhstan imported. This shows that in 2023 a part of the cars, and probably more expensive ones, passed through Kazakhstan to Russia, which can make the average value of imports in 2023 higher than the value of cars that stayed in the country.

4. Discussion

The main result of this study is that the large fall in the average value of imported cars in Kazakhstan's statistics is mostly not a fall in car prices. More than half of the fall from 2023 to 2025 (58%) disappears when the light units are removed. These units are most likely light vehicles that China records as motorcycles and Kazakhstan records as small passenger cars. For a person who downloads only the total of heading 8703, they look like tens of thousands of very cheap Chinese cars.

The rest of the fall has two different sources. For Chinese cars, it is real. Both Kazakhstan and China show that Chinese cars of the same type became cheaper, and the largest fall was in plug-in hybrids and electric cars. At the same time, more of these more expensive types were imported, so the total effect for China was a moderate fall of 12%. This supports the idea that Chinese cars got cheaper, but only for a part of the change.

For Korean and German cars the fall exists only in Kazakhstan's records. Several explanations are possible, and Comtrade data cannot separate them. The first is the difference between the country of origin and the country of shipment. Kazakhstan records German cars by origin, so a used German car bought in another country is recorded as German but does not

appear in Germany's exports to Kazakhstan. This fits the fact that Kazakhstan recorded twice as many German cars as Germany reported. The second is the age of the cars. Since December 2024, importers who are not official dealers can bring only cars older than three years [7], so more used cars with lower values may have been imported in 2025. However, the average Korean car had already fallen from 13,204 to 7,263 dollars in 2024, before this rule, so the rule cannot explain everything. The third is undervaluation, when the value written in the declaration is lower than the real price. Mirror gaps of this size are a common warning sign of undervaluation [4, 5], but they are not proof, because the cars in the two records may not be the same cars.

The difference in the number of Korean cars is also interesting. Korea reported four times more cars sent to Kazakhstan than Kazakhstan recorded. One possible explanation is that some cars sent to Kazakhstan continued to other countries of the Eurasian Economic Union, which is the same route that Chupilkin, Javorcik and Plekhanov describe for goods going to Russia through Central Asia [6]. Kazakhstan's own exports to Russia in 2023 also show this route. Another possibility is the "grey" import of cars that are first registered in another country of the Union and then brought to Kazakhstan [7].

From my experience with declarations, I know that each declaration contains the year of manufacture and the exact model, which Comtrade does not show. With declaration-level data it would be possible to check directly whether the Korean and German cars were old used cars or new cars with low declared values. This would be the next step after this study.

5. Limitations

This study uses only official trade statistics, not the prices of cars in Kazakhstan's shops. Unit values describe what is written in customs records, and they can be affected by the classification and valuation of goods. Kazakhstan's data before 2023 are incomplete, so the long-term change in China's share is described only roughly. The idea that the light units are vehicles recorded as motorcycles by China is based on the closeness of the numbers, the weight and the value, not on direct evidence. Some of Kazakhstan's quantities for 2025 are marked in Comtrade as estimated. Mirror comparisons are affected by the difference between CIF and FOB values and between the country of origin and the country of destination. There is no mirror data for Russia after 2021. Finally, the shift-share decomposition depends on the engine groups, and a finer split, for example by brand, could change the size of the mix effect.

6. Conclusion

Between 2015 and 2025 China became the largest source of passenger cars imported into Kazakhstan, with 56.5% of cars and 49.3% of their value in 2025 when light units are excluded. The average value of an imported unit fell by 39% between 2023 and 2025, but most of this fall came from light units that are most likely not passenger cars. For real cars the fall was 16%. Chinese cars became cheaper by about 12% in Kazakhstan's records and 17% in China's, so this part of the story is real, but the large falls for Korean and German cars exist only in Kazakhstan's records and not in the exporters'. Anyone who uses these statistics should look at the subheadings and compare them with mirror data before drawing conclusions about prices.

Author contribution and use of AI

The author chose the topic from his own experience with customs declarations, wrote the research plan and the main research question, and wrote the first Python code to download and process the Comtrade data for one year. The author's mentor extended the data to all years, partners and subheadings, checked the data, carried out the analysis, made the figures, and helped write and edit the English text. Generative AI tools were used to help with editing. Every

number in the paper was calculated from the Comtrade data by the analysis code and checked by the author's mentor.

Data availability

All data come from the public UN Comtrade database. The Python code that downloads the data and produces every number, table and figure in this paper is available from the author on request.

References:

1. United Nations. UN Comtrade Database: annual trade data by commodity, Harmonized System, Kazakhstan and partners, 2015-2025. URL: https://comtradeplus.un.org (accessed: 02.10.2026).

2. Schott P.K. Across-product versus within-product specialization in international trade // The Quarterly Journal of Economics. 2004. Vol. 119, No. 2. P. 647-678. DOI: 10.1162/0033553041382201.

3. Hallak J.C. Product quality and the direction of trade // Journal of International Economics. 2006. Vol. 68, No. 1. P. 238-265. DOI: 10.1016/j.jinteco.2005.04.001.

4. Bhagwati J. On the underinvoicing of imports // Bulletin of the Oxford University Institute of Economics & Statistics. 1964. Vol. 27, No. 4. P. 389-397. DOI: 10.1111/j.1468-0084.1964.mp27004007.x.

5. Raballand G., Cantens T., Arenas G. Mirror trade statistics: a tool to help identify customs fraud // Reform by Numbers: Measurement Applied to Customs and Tax Administrations in Developing Countries / ed. by T. Cantens, R. Ireland, G. Raballand. Washington, DC: World Bank, 2012. P. 103-120. DOI: 10.1596/9780821397138_ch06.

6. Chupilkin M., Javorcik B., Plekhanov A. The Eurasian roundabout: trade flows into Russia through the Caucasus and Central Asia // European Economic Review. 2026. Vol. 187. Art. 105340. DOI: 10.1016/j.euroecorev.2026.105340.

7. The Times of Central Asia. Why Kazakhstan is tightening requirements for importing foreign cars. 4 February 2025. URL: https://timesca.com/why-kazakhstan-is-tightening-requirements-for-importing-foreign-cars/ (accessed: 02.10.2026).

8. World Customs Organization. Harmonized Commodity Description and Coding System, 2022 edition: heading 87.03. Brussels: WCO, 2022. URL: https://www.wcoomd.org/en/topics/nomenclature/instrument-and-tools/hs-nomenclature-2022-edition.aspx (accessed: 02.10.2026).

УДК 338.262.4

Салимбаев Ербол Аманжолович
MBA, Докторант DBA,
Алматы Менеджмент Университет,
г. Алматы (Казахстан)

АНАЛИЗ ИНВЕСТИЦИОННОЙ АКТИВНОСТИ КАК ИНДИКАТОР ИЗМЕНЕНИЙ ПОТЕНЦИАЛЬНОГО СПРОСА НА КОНСАЛТИНГОВЫЕ УСЛУГИ

Аннотация: Данная статья написана с целью проведения анализа инвестиционной активности в промышленность, строительство и транспортно-логистическую инфраструктуру для понимания объема потенциального спроса на консалтинговые услуги компании. В качестве методов выбраны следующие: анализ, синтез, сравнение и обобщение. Результатом исследования выступает анализ статистических данных, который свидетельствует о сохранении долгосрочной положительной динамики инвестиционной активности в экономике Республики Казахстан. В качестве вывода служит обобщение о том, что выявленные тенденции формируют благоприятные предпосылки для развития рынка консалтинговых услуг.

Ключевые слова: анализ, инвестиционная активность, консалтинговая компания.

Ускорение темпов роста строительной отрасли формирует предпосылки не только для расширения долгосрочного рынка услуг строительного контроля, управления проектами и экспертного сопровождения инвестиционно-строительной деятельности, но и для увеличения текущего спроса на консалтинговые услуги.

Настоящая статья выполнена на основе деятельности консалтинговой компании, значительная часть заказчиков которой находит свое применение в инвестиционно-строительной сфере. В этой связи динамика инвестиций в основной капитал может рассматриваться как один из ключевых внешних индикаторов, определяющих объем потенциального спроса на консалтинговые услуги компании.

Наибольший интерес для настоящего исследования представляют инвестиции в промышленность, строительство и транспортно-логистическую инфраструктуру, поскольку именно в данных секторах экономики формируется основной спрос на услуги строительно-технической экспертизы, финансово-технического контроля, проектного сопровождения и иных консультационных услуг, оказываемых исследуемой компанией. В связи с этим анализ инвестиционной активности в указанных отраслях позволяет оценить состояние внешней среды организации, выявить тенденции изменения потенциального спроса на консалтинговые услуги и определить перспективы дальнейшего развития исследуемого предприятия.

Кроме того, изменение инвестиционной активности в указанных отраслях может рассматриваться в качестве одного из ключевых индикаторов внешних рисков и возможностей, оказывающих влияние на устойчивость консалтинговой компании.

В целях повышения наглядности анализа отдельные отрасли были сгруппированы в зависимости от масштабов инвестиционной активности. Такое разделение обусловлено значительным различием абсолютных объемов инвестиций между секторами экономики,

что затрудняет визуальное восприятие динамики и структурных изменений при построении единых графических материалов.

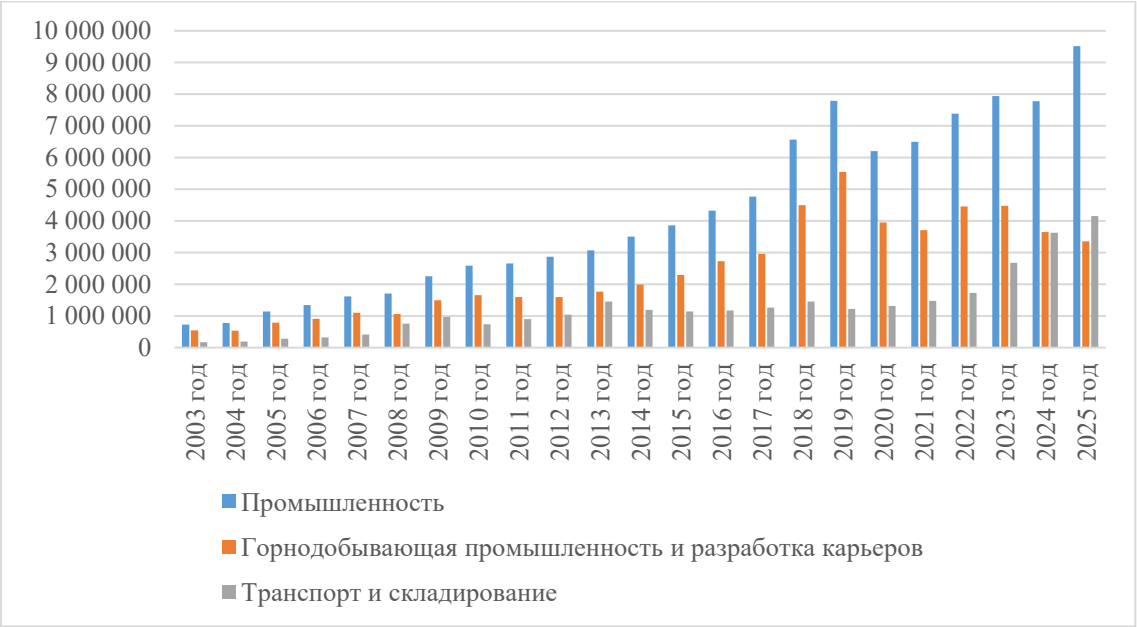


Рисунок 1. Инвестиции в основной капитал по направлениям использования 2003-2025 гг. (промышленность, горнодобывающая промышленность и разработка карьеров, транспорт и складирование), млн. тенге

Примечание - составлено автором на основе данных источника [1]

Анализ отраслевой структуры инвестиций показывает, что промышленность на протяжении всего рассматриваемого периода остается одним из крупнейших направлений вложения капитала в экономике Казахстана. Объем инвестиций в промышленность увеличился с 724,8 млрд тенге в 2003 году до 9,5 трлн тенге в 2025 году, то есть более, чем в 13 раз. Несмотря на отдельные периоды замедления инвестиционной активности, связанные с мировыми экономическими кризисами и колебаниями сырьевых рынков, промышленный сектор сохраняет устойчивое значение для экономики страны.

В структуре промышленных инвестиций доминирующее положение занимает горнодобывающая промышленность. Вместе с тем после 2019 года наблюдается тенденция снижения объемов инвестиций в данный сектор: с 5,6 трлн тенге в 2019 году до 3,3 трлн тенге в 2025 году. Данная тенденция может свидетельствовать о постепенной диверсификации экономики, изменении инвестиционных приоритетов и росте вложений в другие отрасли экономики. Для консалтинговых компаний подобные изменения означают необходимость расширения отраслевой специализации и поиска новых сегментов рынка.

Особый интерес представляет динамика инвестиций в транспорт и складирование. За период с 2003 по 2025 годы объем инвестиций в транспортно-логистическую инфраструктуру увеличился почти в 24 раза – с 174,1 млрд тенге до 4,2 трлн тенге. Особенно высокие темпы роста наблюдаются после 2020 года, что обусловлено активным развитием транзитного потенциала Казахстана, реализацией инфраструктурных проектов и усилением роли страны в международных транспортных коридорах.

Рост инвестиций в промышленность и транспортно-логистическую инфраструктуру оказывает непосредственное влияние на развитие рынка консалтинговых услуг. Реализация крупных инвестиционных проектов сопровождается увеличением потребности в услугах технического и финансово-технического контроля, экспертизы, управления проектами и оценки рисков. Следовательно, выявленные тенденции формируют благоприятные условия для расширения спроса на услуги консалтинговой компании и создают дополнительные возможности для повышения ее устойчивости.

Таким образом, анализ инвестиционной активности показывает, что внешняя среда характеризуется сохранением значительного инвестиционного потенциала в выбранных секторах. Несмотря на отдельные структурные изменения и отраслевые колебания, наблюдается долгосрочная тенденция роста инвестиций в промышленность, строительство и транспортно-логистическую инфраструктуру, что создает предпосылки для дальнейшего увеличения спроса на консалтинговые услуги в области строительного контроля, экспертизы и сопровождения инвестиционных проектов.



Рисунок 2. Инвестиции в основной капитал по направлениям использования 2003-2025 гг., млн. тенге

Примечание - составлено автором на основе данных источника [1]

Анализ инвестиционной активности показывает, что в последние годы в Казахстане наблюдается устойчивый рост вложений в обрабатывающую промышленность. Если в 2003 году объем инвестиций в данный сектор составлял 123,6 млрд тенге, то к 2025 году он увеличился до 2,98 трлн тенге, то есть более, чем в 24 раза. Особенно высокие темпы роста наблюдаются после 2020 года, когда объем инвестиций увеличился с 1,08 трлн тенге до почти 3 трлн тенге. Данная тенденция свидетельствует об усилении государственной политики диверсификации экономики, развитии несырьевого сектора и модернизации производственных мощностей.

Рост инвестиций в обрабатывающую промышленность сопровождается реализацией новых промышленных проектов, строительством производственных объектов и модернизацией существующей инфраструктуры. Это формирует дополнительный спрос на услуги строительно-технической экспертизы, финансово-

технического контроля, оценки рисков и сопровождения инвестиционных проектов, что оказывает положительное влияние на развитие рынка консалтинговых услуг.

Существенный рост также наблюдается в сфере снабжения электроэнергией, газом, паром и горячей водой. За период с 2003 по 2025 годы объем инвестиций увеличился с 39 млрд тенге до 2,23 трлн тенге, то есть более, чем в 57 раз. Наиболее интенсивный рост инвестиций отмечается после 2019 года, что обусловлено необходимостью модернизации энергетической инфраструктуры, строительством новых генерирующих мощностей и увеличением потребностей экономики в энергетических ресурсах.

Развитие энергетической инфраструктуры имеет непосредственное значение для консалтинговых компаний, поскольку реализация крупных энергетических проектов требует осуществления технического надзора, экспертного сопровождения, финансового контроля и управления проектными рисками. Следовательно, рост инвестиционной активности в обрабатывающей промышленности и энергетическом секторе формирует благоприятные внешние условия для расширения спроса на услуги и способствует повышению устойчивости компании.

Таким образом, увеличение инвестиционной активности в несырьевых отраслях экономики свидетельствует о структурной трансформации экономики Казахстана и расширении потенциального рынка консалтинговых услуг, связанных с сопровождением инвестиционно-строительных проектов, техническим контролем и экспертизой.

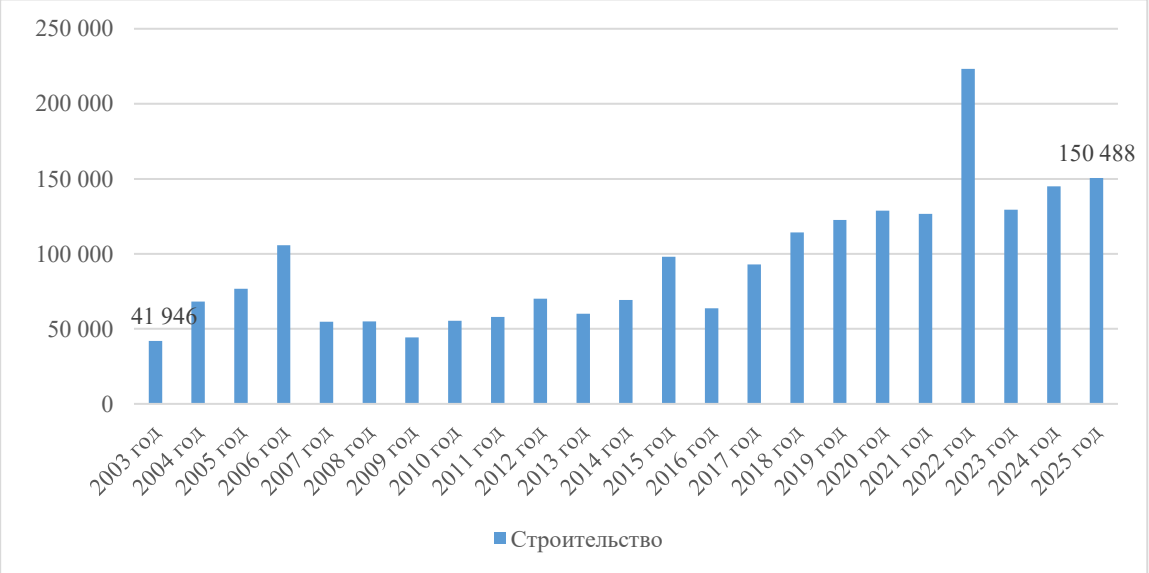


Рисунок 3. Инвестиции в основной капитал по направлениям использования 2003-2025 гг., млн. тенге

Примечание - составлено автором на основе данных источника [1]

Анализ инвестиций в строительную отрасль показывает, что данный сектор характеризуется высокой волатильностью и значительной зависимостью от общеэкономической конъюнктуры. В период 2003-2006 гг. наблюдался рост объемов инвестиций в строительство с 41,9 млрд тенге до 105,8 млрд тенге, что было обусловлено активным развитием рынка недвижимости и высоким уровнем инвестиционной активности в экономике страны. Однако уже в 2007-2009 гг. произошло существенное сокращение инвестиций, объем которых снизился более, чем в два раза и составил 44,2 млрд тенге в 2009 году. Основными причинами стали мировой финансовый кризис,

сокращение доступности кредитных ресурсов и снижение деловой активности в строительном секторе.

В последующие годы инвестиции в строительство демонстрировали неустойчивую динамику. Несмотря на отдельные периоды роста, объем инвестиций оставался чувствительным к изменениям макроэкономической ситуации. Наиболее значительное увеличение произошло в 2022 году, когда объем инвестиций достиг 223,2 млрд тенге, что более чем в 1,7 раза превышало показатель предыдущего года. Данный рост был обусловлен реализацией казахстанских государственных инфраструктурных программ, активизацией жилищного строительства и увеличением бюджетных вложений в развитие инженерной и социальной инфраструктуры.

По состоянию на 2025 год объем инвестиций в строительство составил 150,5 млрд тенге, что свидетельствует о сохранении относительно высокого уровня инвестиционной активности в отрасли. Для консалтинговых компаний строительный сектор имеет особое значение, поскольку именно в данной сфере формируется значительный спрос на услуги строительного контроля, технической экспертизы, финансово-технического аудита и сопровождения инвестиционно-строительных проектов.

Таким образом, строительная отрасль характеризуется циклическим характером развития и высокой чувствительностью к кризисным явлениям, что одновременно создает как дополнительные возможности для роста спроса на консалтинговые услуги в периоды инвестиционного подъема, так и повышенные риски сокращения объемов заказов в периоды экономического спада. В связи с этим состояние строительного сектора следует рассматривать как один из ключевых внешних факторов, определяющих устойчивость деятельности консалтинговой компании.



Рисунок 4. Изменение структуры инвестиций в основной капитал по направлениям использования 2003-2025 гг., %

Примечание - составлено автором на основе данных источника [1]

Анализ отраслевой структуры инвестиций в основной капитал показывает существенные изменения в направлениях инвестиционной активности Республики Казахстан за период 2003-2025 гг. В 2003 году более половины всех инвестиций приходилось на промышленность (54,6 %), из которых 41,2 % были сосредоточены в

горнодобывающем секторе. Это также свидетельствует о высокой сырьевой направленности экономики и значительной зависимости инвестиционных процессов от добывающих отраслей.

Уже к 2025 году наблюдается заметное изменение структуры инвестиций. Несмотря на сохранение доминирующей роли промышленности (40,5 %), доля горнодобывающей отрасли сократилась более, чем в три раза и составила лишь 14,3 %. Одновременно значительно возросла доля обрабатывающей промышленности, увеличившись с 9,3 % до 12,7 %, а также существенно вырос удельный вес инвестиций в транспорт и складирование – с 13,1 % до 17,7 %. Особенно заметным является рост доли инвестиций в энергетическую инфраструктуру, которая увеличилась с 2,9 % до 9,5 %.

Выявленные изменения свидетельствуют о постепенной диверсификации экономики Казахстана, снижении зависимости от сырьевого сектора и усилении инвестиционной активности в инфраструктурных и производственных отраслях. Для консалтинговых компаний данные тенденции имеют положительное значение, поскольку реализация проектов в транспортной, энергетической и обрабатывающей сферах сопровождается повышенным спросом на услуги строительно-технической экспертизы, финансово-технического контроля, управления проектами и оценки инвестиционных рисков.

Таким образом, структурные изменения инвестиционной активности создают благоприятные предпосылки для расширения рынка консалтинговых услуг и одновременно способствуют снижению отраслевой зависимости консалтинговой компании от отдельных сегментов строительного рынка, повышая возможности диверсификации клиентского портфеля и устойчивости компании в условиях кризисных явлений.

Экономика Казахстана постепенно переходит от преимущественно сырьевой модели инвестиционного развития к более диверсифицированной структуре, характеризующейся ростом роли инфраструктурных и обрабатывающих отраслей.

Для составления наиболее полной картины характеристики общей инвестиционной активности на территории РК важно рассмотреть еще и сектор жилищного строительства.

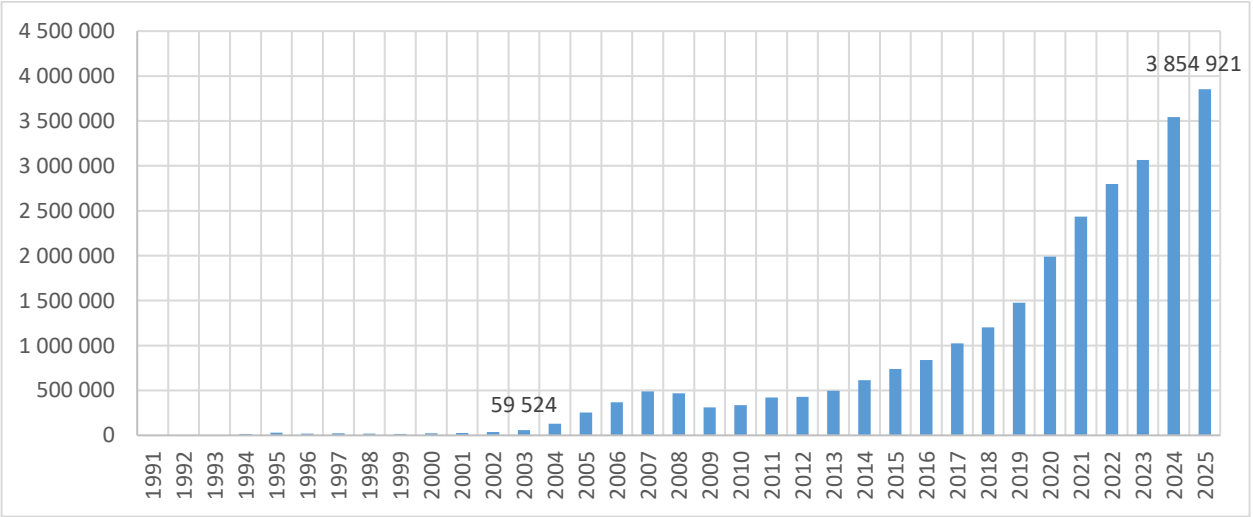


Рисунок 5. Движение объема инвестиций в жилищное строительство РК, млн. тенге

Примечание - составлено автором на основе данных источника [1]

Анализ инвестиций в жилищное строительство Республики Казахстан показывает наличие устойчивой тенденции роста на протяжении последних двух десятилетий. Если в 2003 году объем инвестиций в жилищное строительство составлял 59,5 млрд тенге, к 2025 году данный показатель увеличился до 3,85 трлн тенге, то есть более, чем в 64 раза. Подобная динамика свидетельствует о значительном повышении роли жилищного строительства в экономике страны и росте инвестиционной привлекательности данного сектора.

Наиболее высокие темпы роста наблюдались в период 2004-2007 гг., когда объем инвестиций увеличился с 130,5 млрд тенге до 490,4 млрд тенге. Однако мировой финансовый кризис 2008-2009 гг. оказал существенное негативное влияние на строительный сектор, вследствие чего инвестиции сократились до 310,8 млрд тенге в 2009 году. Несмотря на кризисные явления, начиная с 2010 года наблюдается постепенное восстановление инвестиционной активности.

Особенно интенсивный рост инвестиций отмечается после 2017 года. За период с 2017 по 2025 годы объем инвестиций в жилищное строительство увеличился с 1,02 трлн тенге до 3,85 трлн тенге, то есть почти в четыре раза. Данная тенденция обусловлена реализацией государственных жилищных программ, развитием ипотечного кредитования, ростом численности населения и повышением спроса на жилую недвижимость.

Рост инвестиционной активности в жилищном строительстве оказывает непосредственное влияние на рынок консалтинговых услуг. Увеличение объемов строительства жилых объектов сопровождается ростом потребности в услугах технического надзора, строительно-технической экспертизы, финансово-технического контроля и сопровождения инвестиционно-строительных проектов. Исходя из этого, становится очевидным, что развитие жилищного строительства формирует благоприятные внешние условия для расширения спроса на консалтинговые услуги в данной сфере и оказывает положительное влияние на устойчивость деятельности компании.

Таким образом, жилищное строительство является одним из наиболее динамично развивающихся сегментов инвестиционной деятельности в Казахстане и выступает важным источником формирования спроса на консалтинговые услуги в сфере строительного контроля и экспертизы.

Проведенный анализ статистических данных свидетельствует о сохранении долгосрочной положительной динамики инвестиционной активности в экономике Республики Казахстан. За последние два десятилетия наблюдается существенное увеличение объемов инвестиций в основной капитал, развитие промышленности, транспортно-логистической инфраструктуры, энергетического комплекса и жилищного строительства. Одновременно происходит постепенное изменение отраслевой структуры инвестиций, выражающееся в снижении относительной роли сырьевого сектора и усилении инвестиционной активности в обрабатывающей промышленности и инфраструктурных отраслях экономики.

Выявленные тенденции формируют благоприятные предпосылки для развития рынка консалтинговых услуг. Реализация крупных инвестиционных и строительных проектов влечет за собой рост потребности в услугах строительно-технической экспертизы, финансово-технического контроля, оценки рисков, сопровождения инвестиционных проектов и иных консультационных услуг. Следовательно, внешняя

макроэкономическая среда в целом характеризуется наличием значительного потенциала для дальнейшего развития компаний, осуществляющих деятельность в сфере инвестиционно-строительного консалтинга.

Список литературы:

1. Бюро национальной статистики Агентства по стратегическому планированию и реформам Республики Казахстан / https://stat.gov.kz/ru/

UDC 334.722

Rambergenov Ramil
Grade 11 student
Academic advisor: Fatima Matyokubova
Technical College
(Bustan, Uzbekistan)

REGIONAL NEW-FIRM FORMATION IN UZBEKISTAN AFTER THE 2017 REFORMS: PERSISTENCE IN SCALE, CONVERGENCE IN INTENSITY, AND RECOMPOSITION TOWARD TRADE

Abstract: Uzbekistan's 2017 economic liberalisation was followed by a roughly threefold increase in new-business registration, but whether that national surge reshaped the regional structure of entrepreneurship has not been tested. This study builds a 224-observation region-year panel (14 regions, 2010-2025) from the National Statistics Committee's Unified State Register of Enterprises and Organisations and estimates beta-convergence on two pre-registered outcomes: new firms per year (levels) and new firms per 1,000 registered enterprises (formation intensity). On the pre-registered intensity outcome, convergence is robust: beta = -0.990 (HC3 SE 0.367, p = 0.007, R-squared = 0.537), negative in 14 of 14 leave-one-out fits. On the levels outcome the estimate is negative but inconclusive-underpowered (beta = -0.235, p = 0.053), with a minimum detectable effect of 0.311 that exceeds the observed coefficient. Rank-stability analysis resolves the apparent tension: the absolute ordering of regions is frozen (Spearman rho = +0.93), while the intensity ordering reshuffles completely (rho = -0.02) and its dispersion compresses. Sectoral decomposition shows the boom was compositionally narrow: wholesale and retail trade gained 9.07 percentage points of share and accounts for 41.5 per cent of the national surge, while construction, manufacturing and finance lost share. The post-2017 period is therefore best described as persistence in scale, convergence in intensity, and recomposition toward trade. No causal attribution to the reforms is claimed.

Keywords: new firm formation; regional convergence; Uzbekistan; entrepreneurship; economic reform

1. Introduction

Where new firms are created, and whether that geography changes when policy changes, is one of the durable questions in regional economics. The stakes are practical: if entrepreneurship concentrates permanently in a capital region, liberalisation can raise national firm counts while leaving the regional structure of opportunity untouched. The empirical literature on mature economies suggests exactly this kind of stickiness. Armington and Acs [2] show that regional variation in new firm formation is explained by industrial density, population growth and income growth, all of which are slow-moving. Andersson and Koster [1] go further, documenting that regional start-up rates in Sweden are persistent and path-dependent, and that regions with high start-up rates exhibit the strongest persistence of all.

Uzbekistan after 2017 is an unusually clean setting in which to ask whether a large policy shock disturbs that structure. Following the change of government, the country undertook a broad liberalisation programme including currency convertibility, tax simplification and a reduction in licensing and inspection burdens. National new-business registration rose from 32,747 in 2016 to 96,743 in 2019, and has since plateaued in the 80,000-100,000 range. A shock

of that size is rare, and it offers a natural test: did the surge redistribute entrepreneurship across the country's regions, or did it simply scale up the existing hierarchy?

The literature offers a theoretical reason to expect institutional reform to matter. Carbonara, Santarelli and Tran [4], using instrumental-variable GMM across 115 countries, find that constitutional provisions protecting the right to conduct business are positively associated with new business density. Tonoyan, Strohmeyer, Habib and Perlitz [11] describe the institutional baseline such reforms act against in transition economies, where weak legal enforcement and closed networks sustain corrupt dealing. Reform of the kind Uzbekistan undertook is therefore a plausible driver of entry, even though establishing that causally is beyond what the present data can support.

What does not exist is a quantitative regional treatment of the Uzbek case. Three close-neighbour studies define the frontier. Tadjiev, Ataev, Akhmetshin and Vasilev [10] assess the effectiveness of Uzbekistan's entrepreneurship-support reforms, but at national and institutional level with no panel estimation. Nizaeva [9] studies regulatory and institutional drivers of small-firm growth across Central Asia using Enterprise Survey data, but the design is a four-country firm-level cross-section whose outcome is employment growth rather than entry, and it contains no within-Uzbekistan regional dimension. Asadov and Bozorov [3] analyse the 2017-2025 reform period in Russian-language policy-narrative form, without firm-entry estimation. A search across English and Russian-language sources returns no published work that estimates new-firm formation across Uzbekistan's regions over time using the official business register.

One boundary on the question should be fixed at the outset. This study measures entry, not durability. Huggins, Prokop and Thompson [5] show that human capital, growth motivation and local conditions explain firm survival within regions, which is a distinct outcome from formation; a region that creates many firms is not thereby shown to sustain them. Nothing below speaks to survival.

This study fills that gap. It contributes, first, a 224-observation region-year panel covering all 14 regions from 2010 to 2025, built from the national business register; second, a pre-registered test of beta-convergence on two distinct outcomes, formation levels and formation intensity, which turn out to behave very differently; third, a rank-stability analysis that reconciles those two results; and fourth, a sectoral decomposition of the national surge that has no counterpart in the existing literature. The analysis plan was fixed in writing before any confirmatory result was produced.

The paper proceeds as follows. Section 2 describes the data, the pre-registered design and the validity checks. Section 3 reports the descriptive series, the convergence estimates, the rank-stability results and the sectoral decomposition. Section 4 interprets these against the persistence literature and states the study's limits. Section 5 concludes.

2. Data and method

The study asks how new-business formation evolved across Uzbekistan's 14 regions following the 2017 reforms, whether initially lagging regions caught up, and how the sectoral composition of new firms shifted over the same period. It is designed to describe and test patterns of regional convergence, not to establish that the reforms caused them.

Data come from the Unified State Register of Enterprises and Organisations (USREO) maintained by the National Statistics Committee of the Republic of Uzbekistan, accessed through the committee's own SDMX interface. Three region-by-year series were retrieved (newly created enterprises, newly created small businesses, and registered small businesses, 2010-2025) together with three national sector-by-year series classified by NACE section

(2013-2025). One scope constraint should be stated plainly: the register publishes new-business creation by region and by sector, but not as a combined region-by-sector cross-tabulation. The two dimensions are therefore analysed as marginals, with the region-by-time panel as the spine of the study and the national sector-by-time series as a second lens. Regional coverage comprises the 12 provinces, the Republic of Karakalpakstan and the city of Tashkent, referred to throughout as the 14 regions.

The design is observational and descriptive-comparative, built around a single national policy episode in 2017. The unit of analysis is the region-year. The panel covers 14 regions over 16 years (2010-2025), giving 224 observations. Two outcomes were pre-registered. The first is the count of newly created enterprises per region per year, used in levels and in logs. The second is the formation rate, defined as newly created enterprises per 1,000 registered enterprises in the same region and year, which normalises for regional size and is the standard start-up-rate outcome in the regional entrepreneurship literature. The principal estimator is a beta-convergence regression in which the growth of formation between the pre-reform period (2013-2016) and the post-reform period (2018-2024) is regressed on the initial pre-reform level, separately for each outcome. A negative coefficient indicates that regions starting lower grew faster. Supporting specifications are a two-way fixed-effects panel model of log formation on a post-2017 indicator with region and year fixed effects, an event study around 2017, a rank-stability analysis, and a national sectoral decomposition. Because the design has only 14 clusters, the fixed-effects specification is tested with a wild-cluster bootstrap rather than asymptotic clustered standard errors.

The pre-registered hypothesis is beta-convergence: the null is that the coefficient on the initial level is zero, and the alternative is that it is negative, meaning initially lower regions grew faster. The hypothesis was stated for both outcomes before estimation. A null result was pre-committed to be reported as a bounded or inconclusive finding according to the design's power, never as evidence of no effect.

All analysis was carried out in Python using pandas for data assembly, statsmodels for estimation, and scipy for rank statistics. Beta-convergence models are estimated by ordinary least squares with HC3 heteroskedasticity-robust standard errors, which are appropriate at this sample size. The fixed-effects specification uses a wild-cluster bootstrap with 9,999 Rademacher replications and the null imposed. Three analysis scripts and seven derived datasets are retained: they regenerate every number reported below.

The panel was assembled from the raw register extracts and verified to contain 14 regions across 16 years with no missing cells. Pre-period (2013-2016) and post-period (2018-2024) means were computed per region for each outcome, and growth was calculated as the log ratio of post to pre. The transition year 2017 is excluded from both windows so that the comparison is not contaminated by the partial-year effect of reforms enacted during it. Before interpreting any coefficient, four validity checks were run: a Jarque-Bera test of residual normality, a Breusch-Pagan test for heteroskedasticity, a full influence diagnostic reporting Cook's distance and leverage for every region, and a leave-one-out re-estimation dropping each region in turn. A minimum detectable effect at 80 per cent power was computed for each specification so that any null could be classified as bounded or as underpowered rather than being read as absence of effect.

This study uses only aggregate, publicly published official statistics. No human participants were involved, no survey or interview was conducted, and no personal or identifying data of any kind was accessed or processed. Informed consent was therefore not

applicable. All source data are openly available from the National Statistics Committee of the Republic of Uzbekistan, and all analysis code and derived datasets are retained so that every reported figure can be reproduced.

3. Results

The national series establishes the scale of the episode. New-enterprise creation was roughly flat between 2010 and 2016, fluctuating between 21,890 and 34,326 per year. It then rose sharply: 41,013 in 2017, 55,011 in 2018 and 96,743 in 2019, a 2.95-fold increase over the 2016 base. The series has since settled onto an 80,000-100,000 plateau. All 14 regions grew; none contracted.

Table 1 – Newly created enterprises in Uzbekistan, national totals by year, 2010–2025

Year	New enterprises
2010	34,326
2011	32,715
2012	21,890
2013	27,819
2014	26,933
2015	27,915
2016	32,747
2017	41,013
2018	55,011
2019	96,743
2020	95,311
2021	102,804
2022	93,611
2023	88,787
2024	79,667
2025	87,775

Table 2 – Regional formation before and after the reforms (period means)

Region	Pre 2013–16	Post 2018–24	Pre rate	Post rate
Andijan	2,143	5,830	96.9	186.3
Bukhara	1,683	5,409	136.7	239.4
Fergana	2,577	7,491	145.2	219.4
Jizzakh	1,203	4,138	141.9	242.6
Karakalpakstan	1,327	4,245	129.0	240.8
Kashkadarya	1,719	6,270	93.7	247.0
Khorezm	1,365	5,166	129.7	261.7
Namangan	1,696	5,213	113.2	212.1
Navoi	822	4,298	121.4	298.7
Samarkand	2,116	8,368	146.8	264.6
Surkhandarya	1,459	4,881	134.1	269.6
Syrdarya	964	2,358	138.9	195.5
Tashkent city	7,066	16,145	168.5	196.7
Tashkent region	2,717	7,609	130.4	193.4

Validity checks pass. Residual normality is not rejected (Jarque-Bera = 1.31, p = 0.519) and homoskedasticity is not rejected (Breusch-Pagan LM = 3.22, p = 0.073); HC3 standard errors are used regardless. The influence diagnostic identifies three regions that carry disproportionate weight in the levels specification: Syrdarya (Cook's D = 0.689), Navoi (0.584) and the city of Tashkent (leverage = 0.606). This matters for interpretation and is reported rather than removed.

On the levels outcome, the convergence coefficient is negative but does not clear conventional significance: beta = -0.235, HC3 SE = 0.122, t = -1.93, p = 0.053, R-squared = 0.345, N = 14. The leave-one-out analysis shows the sign is stable, negative in 14 of 14 fits with beta ranging from -0.319 to -0.151, but significance is not: p is below 0.05 in only 2 of 14 fits, and the result swings from p = 0.436 when the city of Tashkent is dropped to p = 0.003 when Syrdarya is dropped. The design's minimum detectable effect at 80 per cent power is 0.311, larger in absolute value than the observed coefficient of 0.235. The levels result is therefore classified as inconclusive and underpowered. It is not evidence that convergence in levels is absent.

On the pre-registered formation-intensity outcome the picture is different and much sharper. Regressing growth in new firms per 1,000 registered enterprises on the initial log intensity gives beta = -0.990, HC3 SE = 0.367, t = -2.70, p = 0.007, R-squared = 0.537, N = 14. Leave-one-out re-estimation returns a negative coefficient in 14 of 14 fits, significant at the 5 per cent level in 13 of 14, with beta ranging from -1.265 to -0.801. This is the outcome that corresponds to the start-up-rate measure used in the benchmark regional literature, and on it convergence is robust.

Table 3 – Beta-convergence estimates, pre-registered outcomes (N = 14)

Outcome	β	HC3 SE	t	p	R^2
Formation levels	−0.235	0.122	−1.93	0.053	0.345
Formation intensity	−0.990	0.367	−2.70	0.007	0.537

The rank-stability analysis reconciles the two results. On the levels outcome, the Spearman correlation between each region's pre-period and post-period rank is +0.93 (p < 0.001): the absolute ordering of regions is essentially frozen, with the city of Tashkent dominant before and after. On the intensity outcome, the same correlation is -0.02 (p = 0.95), meaning the ordering is completely reshuffled, and the cross-regional dispersion of log intensity compresses from 0.158 to 0.146. Navoi moves from 11th to 1st on formation intensity, Kashkadarya from 14th to 5th, and the city of Tashkent falls from 1st to 11th. Both patterns are real and they describe different things: which regions are biggest, and how intensively each region is creating firms relative to its existing business stock.

Table 4 – Rank stability and dispersion, pre-period versus post-period

Outcome	Spearman ρ	p	σ pre	σ post
Formation levels (new firms per year)	+0.93	<0.001	0.527	0.438

Formation intensity (per 1,000 stock)	−0.02	0.946	0.158	0.146

The two-way fixed-effects interaction between the post-2017 indicator and the initial level is negative but not significant (c = -0.152, wild-cluster p = 0.277 with 9,999 replications and 14 clusters). This is the panel expression of the levels claim, and its null is consistent with the underpowered levels result rather than contradicting the intensity result. An earlier pipeline specification that returned p = 0.000 was a small-cluster artefact and is not reported as a result.

The sectoral decomposition shows the boom was compositionally narrow. Wholesale and retail trade including motor-vehicle repair rose from 26.11 per cent to 35.18 per cent of new firms, a gain of 9.07 percentage points, and alone accounts for 41.51 per cent of the entire national increase in new-firm creation. Construction lost 4.37 percentage points of share, manufacturing lost 4.04, and financial and insurance activities lost 2.18, contracting to 0.32 times its 2016 count in absolute terms. The Herfindahl-Hirschman index of sectoral concentration rose from 1,396 to 1,624: the composition of new firms became more concentrated over the boom, not more diversified.

Table 5 – Sectoral composition of new firms, largest movers, 2016 versus 2024

Sector (NACE section)	2016 (%)	2024 (%)	Change (pp)	Share of surge (%)
Wholesale and retail trade; vehicle repair	26.11	35.18	+9.07	41.51
Agriculture, forestry and fisheries	10.25	11.44	+1.19	12.28
Other service activities	2.23	3.38	+1.14	4.17
Education	1.92	2.93	+1.01	3.63
Information and communication	2.89	3.68	+0.79	4.23
Financial and insurance activities	2.51	0.34	−2.18	−1.18
Manufacturing	19.25	15.22	−4.04	12.40
Construction	9.30	4.93	−4.37	1.88

4. Discussion

The central result is that two standard measures of the same phenomenon tell opposite stories, and that the difference is substantive rather than technical. Measured by how many firms are created, Uzbekistan's regional hierarchy did not move: the pre-period and post-period rankings correlate at 0.93, and the city of Tashkent remained dominant throughout a national tripling. Measured by how intensively each region creates firms relative to its existing stock, the hierarchy inverted: the rank correlation is effectively zero, dispersion narrowed, and regions that began least intensive, notably Navoi and Kashkadarya, ended most intensive.

This refines rather than contradicts the persistence literature. Andersson and Koster [1] established that regional start-up rates are persistent and path-dependent in Sweden. Our levels

result is consistent with that stickiness under a very large shock. But their persistence finding concerns start-up rates, and it is precisely on the rate outcome that we observe robust convergence and a complete reshuffle. The reasonable reading is that persistence in the absolute scale of regional entrepreneurship can coexist with rapid convergence in its intensity, and that a study reporting only one of the two outcomes would draw a confident but incomplete conclusion. Kangasharju [7] is relevant here as well: he documents that while long-run beta-convergence in Finland runs at roughly 2 per cent per year, short-run beta is unstable. Our eight-year window is short by that standard, which is one reason the levels estimate is fragile and the design was checked for power rather than read at face value.

The sectoral result carries a caution that should not be buried. The surge was concentrated in low-capital retail trade, which accounts for over 40 per cent of the national increase. La Porta and Shleifer [8] document that informal firms account for up to half of economic activity in developing countries. A registration boom concentrated in exactly the sector where informal activity is most common may therefore partly reflect the formalisation of pre-existing trade rather than the creation of genuinely new economic activity. The available data cannot distinguish these, and we do not claim to. The finding is that registered entry recomposed toward trade; whether that entry is new or newly visible is an open question and a natural target for follow-up work with firm-level or survey data.

Several limits bound what can be concluded. First, no causal attribution to the 2017 reforms is possible. There is no control group, and the event study shows a non-flat pre-trend with five of six pre-2016 coefficients significant, so the timing of the change cannot be pinned to 2017. Throughout, the language is that the surge coincides with or follows the reforms, never that it was caused by them. Second, the levels outcome is underpowered at N = 14 and is reported as inconclusive; the convergence claim rests on the intensity outcome, which is robust. Third, the modifiable areal unit problem applies directly to this design. Beta-convergence estimates depend on the spatial partition chosen, and 14 administrative regions is a coarse and politically determined partition; a finer district-level partition could give different estimates. Fourth, the rate denominator is the registered stock, which itself grows during the boom. This deflates the measured rate in fast-growing regions and therefore biases against finding rate convergence where growth was strongest, which makes the reported convergence result conservative. Fifth, nothing here resolves at region-by-sector level, because that cross-tabulation is not published.

It is worth noting that stickiness of this kind is not peculiar to transition economies. Iammarino, Rodríguez-Pose and Storper [6] document persistent regional inequality across Europe and argue that regions follow several distinct performance modes rather than converging uniformly. Uzbekistan's frozen levels hierarchy sits comfortably alongside that evidence, which suggests the pattern reflects general features of economic geography rather than anything specific to the reform episode studied here.

Within those limits the implication for policy is straightforward. Reform-era liberalisation in Uzbekistan appears to have broadened participation in firm creation, in the sense that lagging regions caught up on formation intensity, without redistributing where economic mass sits. If the policy objective is regional balance in absolute terms, an intensity-based catch-up is a weaker achievement than headline national growth suggests. The concentration of that catch-up in retail trade, together with the absolute contraction of new financial and manufacturing firms, further suggests that the composition of the boom deserves as much policy attention as its size.

5. Conclusion

This study built the first region-by-time panel of new-firm formation for Uzbekistan and used it to test whether the post-2017 surge changed the country's regional entrepreneurial structure. Three findings emerge. National new-firm creation roughly tripled between 2016 and 2019 and every region participated. On the pre-registered formation-intensity outcome, initially lower-intensity regions converged robustly on the rest (beta = -0.990, p = 0.007, negative in all 14 leave-one-out fits), while the absolute hierarchy of regions remained frozen (rank correlation 0.93). The boom was compositionally narrow, with retail trade contributing over 41 per cent of the national increase while construction, manufacturing and finance all lost share. The period is best summarised as persistence in scale, convergence in intensity, and recomposition toward trade.

The main limitations are the absence of any causal identification, an underpowered levels specification that is reported as inconclusive rather than null, sensitivity of the spatial estimates to the administrative partition used, and the inability to separate genuinely new entry from the formalisation of existing informal trade.

Three extensions follow naturally. A district-level partition would test how far the convergence result depends on the choice of spatial units. Firm-level or survey data would allow new entry to be distinguished from formalisation, which is the single largest interpretive ambiguity in the present results. And a comparative design covering other Central Asian transition economies would place Uzbekistan's episode in context and open the possibility of a genuine counterfactual, which the single-country design here cannot provide.

References:

1. Andersson M., Koster S. Sources of persistence in regional start-up rates: Evidence from Sweden // Journal of Economic Geography. 2011. Vol. 11, № 1. P. 179–201. DOI: 10.1093/jeg/lbp069.

2. Armington C., Acs Z. J. The determinants of regional variation in new firm formation // Regional Studies. 2002. Vol. 36, № 1. P. 33–45. DOI: 10.1080/00343400120099843.

3. Асадов Х. С., Бозоров А. Т. Socio-economic reforms and development outcomes of Uzbekistan in 2017–2025 [in Russian] // Eurasian Studies. 2026. № 2. DOI: 10.48647/icca.2026.39.76.006.

4. Carbonara E., Santarelli E., Tran H. T. De jure determinants of new firm formation: How the pillars of constitutions influence entrepreneurship // Small Business Economics. 2016. Vol. 47, № 1. P. 139–162. DOI: 10.1007/s11187-016-9715-z.

5. Huggins R., Prokop D., Thompson P. Entrepreneurship and the determinants of firm survival within regions: Human capital, growth motivation and locational conditions // Entrepreneurship & Regional Development. 2017. Vol. 29, № 3–4. P. 357–389. DOI: 10.1080/08985626.2016.1271830.

6. Iammarino S., Rodríguez-Pose A., Storper M. Regional inequality in Europe: Evidence, theory and policy implications // Journal of Economic Geography. 2019. Vol. 19, № 2. P. 273–298. DOI: 10.1093/jeg/lby021.

7. Kangasharju A. Beta convergence in Finland: Regional differences in speed of convergence // Applied Economics. 1998. Vol. 30, № 5. P. 679–687. DOI: 10.1080/000368498325660.

8. La Porta R., Shleifer A. Informality and development // Journal of Economic Perspectives. 2014. Vol. 28, № 3. P. 109–126. DOI: 10.1257/jep.28.3.109.

9. Nizaeva M. Regulatory and institutional factors driving small firms' growth in Central Asia // Eurasian Journal of Business and Economics. 2024. Vol. 17, № 34. P. 167–184. DOI: 10.17015/ejbe.2024.034.08.

10. Tadjiev B. U., Ataev J. E., Akhmetshin E. M., Vasilev V. L. Assessment of the effectiveness of the reforms to support entrepreneurship in Uzbekistan // E3S Web of Conferences. 2023. Vol. 449. P. 03002. DOI: 10.1051/e3sconf/202344903002.

11. Tonoyan V., Strohmeyer R., Habib M., Perlitz M. Corruption and entrepreneurship: How formal and informal institutions shape small firm behavior in transition and mature market economies // Entrepreneurship Theory and Practice. 2010. Vol. 34, № 5. P. 803–832. DOI: 10.1111/j.1540-6520.2010.00394.x.

UDC 336.7

Abdesh Dinara
Grade 12 student
Nazarbayev Intellectual School
(Astana, Kazakhstan)

FROM A WOMEN'S LEAD TO A MEN'S LEAD: THE GENDER GAP IN DIGITAL PAYMENTS IN KAZAKHSTAN AND THREE CENTRAL ASIAN NEIGHBOURS, 2021-2024

Abstract: For most of the last decade, women in Kazakhstan were slightly more likely than men to have a financial account. Using the World Bank's Global Findex surveys of 2021 and 2024, I compared women and men in Kazakhstan, Kyrgyzstan, Tajikistan and Uzbekistan in account ownership, digital payments and saving. In Kazakhstan the gap reversed. In 2021, 81.8% of women and 73.6% of men made or received a digital payment; in 2024 the figures were 82.1% and 88.3%. The change in the gap, 14.3 percentage points, is larger than its approximate 95% margin of error of 7.9 points. The reversal happened because men's use grew by 14.7 points while women's stayed almost the same. Accounts and saving moved in the same direction. Two other indicators changed at the same time: the share of men who received wages into an account rose by 22.9 points, and the share of women who received a government transfer or pension fell by 9.4 points. The neighbours followed different paths. In Tajikistan a gap in favour of men opened, in Uzbekistan an earlier male advantage in accounts and saving almost disappeared, and in Kyrgyzstan the gap hardly changed. These are country-level survey figures and cannot show causes, but they suggest that in Kazakhstan the way people receive money, through wages or government transfers, matters for who uses digital finance.

Keywords: gender gap; digital payments; financial inclusion; Global Findex; Kazakhstan; Central Asia; wages

Introduction

In most discussions of financial inclusion, the gender gap means one thing: women are less likely than men to have an account, save formally or use digital payments. Using individual data from 98 developing countries, Demirgüç-Kunt, Klapper and Singer found that gender remains significantly related to the use of financial services even after controlling for income, education, employment, age and rural residence, and that legal restrictions on women's work and inheritance help explain where the gap is largest [1]. A study of 144 countries in the 2017 Findex data also found statistically significant differences between men and women in account ownership, saving and digital payments [2].

Kazakhstan did not fit this picture. In every Findex round from 2011 to 2021, women in Kazakhstan were slightly more likely than men to have an account: 43.7% against 40.2% in 2011, 60.3% against 56.9% in 2017 and 83.6% against 78.3% in 2021 [3]. In digital payments, women were also ahead in 2017 and 2021. The 2024 round is the first in which men are ahead on both measures. Because the overall level is high, 87.0% of adults in Kazakhstan had an account in 2024 [3], a change like this can stay hidden inside the national average.

Research on women's finance in Kazakhstan has mostly looked at credit. Bekturganova and Kurmasheva studied how the structure of the banking sector affects women's access to loans and found that credit for women did not grow in proportion to banking resources [4], and

Kireyeva and Kurmasheva linked women's entrepreneurial lending to regional output and the share of wages in GDP [5]. At the regional level, Poghosyan showed that the probability of having a financial institution account in the Caucasus and Central Asia is lower than in Central and Eastern Europe across gender, income and education groups [6]. What has not been described, as far as I could find, is how the gender gap in everyday digital payments changed between the two latest Findex rounds, and whether Kazakhstan's change is part of a regional pattern.

This study asks three questions:

• How did the gap between men and women in digital payments, accounts and saving in Kazakhstan change between 2021 and 2024, and is the change larger than the survey's margin of error?

• Did the gap change because women's use fell, or because men's use grew?

• Did the same thing happen in Kyrgyzstan, Tajikistan and Uzbekistan?

Data and method

Data. All figures come from the Global Findex Database 2025, which contains the 2021 and 2024 survey rounds [7, 8], and were downloaded from the World Bank's data service on 6 October 2026 [3]. In each economy Gallup interviews about 1,000 randomly selected people aged 15 and older. In all four countries the interviews were face to face in both rounds, so a change in how the survey was carried out cannot explain the results [9, 10]. In Uzbekistan in 2024, the Republic of Karakalpakstan, about 6% of the population, was excluded from the sample [10]. I used only the published percentages for women and men, not individual answers.

Indicators. The main indicator is the share of adults who made or received a digital payment in the past year, which includes paying with a card or phone, paying bills online and receiving wages, transfers or other payments into an account. I also used account ownership, account ownership at a bank or similar financial institution, having saved any money, having saved at a bank, and owning a debit card. To look at what was behind the change in Kazakhstan, I split digital payments into made and received, and used three more indicators: having received wages, having received wages into an account, and having received a government transfer or pension. For Kazakhstan in 2021 mobile money was not measured, so in that year the share with an account is the same as the share with a bank account (83.6% of women and 78.3% of men). To compare like with like across years, I therefore also report the bank account indicator.

Calculations. For each indicator the gap is the share of men minus the share of women, in percentage points, so a negative gap means women are ahead. The change in the gap is the gap in 2024 minus the gap in 2021. All values were calculated from the unrounded figures in the database.

Margins of error. The database does not give standard errors, so I estimated them. The Findex methodology notes give the sample size and the design effect for each country and year, for example 1.46 for Kazakhstan in 2021 and 1.24 in 2024 [9, 10]. I assumed that each sample is split evenly, 500 women and 500 men, and estimated the standard error of each percentage as the square root of the design effect times p(1 - p)/500. The standard error of a gap combines those of the two groups, and the standard error of a change combines those of the two years, because the two rounds are independent samples. I report 95% margins of error as 1.96 standard errors and call a change clear only when it is larger than its margin. These are rough estimates, because the real number of women in each sample is not exactly 500.

Results

Kazakhstan: a women's lead became a men's lead

Table 1 shows the main indicators for Kazakhstan. In 2021 women were ahead or level on every measure, and in 2024 men were ahead on every measure.

Table 1. Women and men in Kazakhstan, Global Findex 2021 and 2024

Indicator (% of adults 15+)	W 2021	M 2021	Gap 2021	W 2024	M 2024	Gap 2024	Change in gap
Made or received a digital payment	81.8	73.6	-8.2	82.1	88.3	+6.2	+14.3 (±7.9)
Account	83.6	78.3	-5.3	85.7	88.5	+2.8	+8.1 (±7.5)
Account at a bank or similar institution	83.6	78.3	-5.3	81.9	86.8	+4.9	+10.2 (±7.7)
Saved any money	42.2	38.0	-4.2	49.5	56.8	+7.2	+11.5 (±10.0)
Saved at a bank or similar institution	19.3	18.9	-0.5	24.4	30.7	+6.3	+6.8 (±8.5)
Owns a debit card	62.2	57.5	-4.6	66.5	69.1	+2.6	+7.2 (±9.8)

Note: W = women, M = men. Gap = men minus women, percentage points; a negative gap means women are ahead. Numbers in brackets are approximate 95% margins of error (see Data and method). Gaps and changes are calculated from unrounded values. Source: Global Findex Database 2025; author's calculations.

In 2021, women in Kazakhstan were ahead of men. The share of women using digital payments was 81.8%, compared with 73.6% of men. This created a gender gap of -8.2 percentage points, meaning that women used digital payments more often. By 2024, the situation had reversed. Men reached 88.3%, while women reached 82.1%, so the gap became +6.2 percentage points in favour of men. This change was mainly caused by the rapid increase among men. The share of men using digital payments increased by 14.7 percentage points, while the share among women increased by only 0.3 percentage points. Therefore, the reversal was not mainly because women used digital payments less. Instead, men caught up and moved clearly ahead.

For digital payments, the gap in each year and the change between the years are all larger than their margins of error: the gap was -8.2 (±6.2) in 2021 and +6.2 (±4.9) in 2024. For accounts the picture is less sharp. In each year taken alone, the account gap is within the margin of error, but the change of 8.1 points is just larger than its margin of 7.5, and with the bank account indicator, which is measured the same way in both years, the change is 10.2 (±7.7). The change in saving any money, 11.5 points, is also just larger than its margin of 10.0. The smaller changes in saving at a bank and debit card ownership go in the same direction but are within the margin of error.

Figure 1 shows that the 2024 result was not a continuation of an old trend. Women had a small lead in account ownership in every round from 2011 to 2021, and their lead in digital payments was largest in 2021. In 2024 both lines crossed.

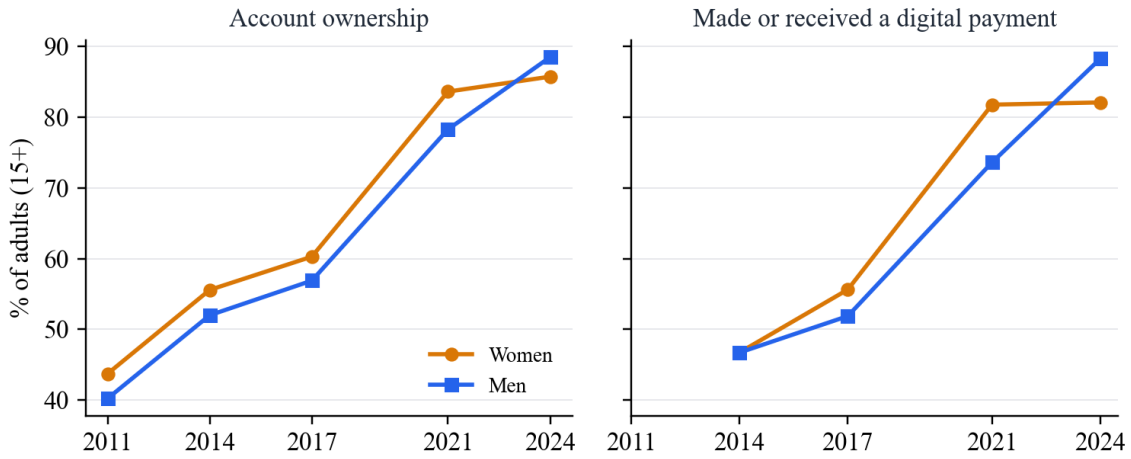


Figure 1. Kazakhstan: account ownership and digital payments, women and men, 2011-2024. Digital payments were not measured in 2011.

Kazakhstan and its neighbours

In Uzbekistan, 2021, men were clearly ahead of women in digital payment use. 48.5% of men used digital payments compared with 35.9% of women, creating a gap of 12.6 percentage points in favour of men. However, by 2024, this difference had become much smaller. Women's digital payment use increased to 48.6%, while men's increased to 51.3%. As a result, the gender gap fell to only 2.7 percentage points. This means that women made much faster progress than men. Women's use increased by 12.7 percentage points, compared with only 2.8 percentage points for men. Therefore, Uzbekistan can be seen as a mirror of Kazakhstan because the countries started with different gender patterns. Kazakhstan moved from a female advantage to a male advantage, while Uzbekistan moved from a strong male advantage towards almost equal use.

The results for Kyrgyzstan and Tajikistan show different trends. In Kyrgyzstan, the gender gap changed only slightly, from +2.9 percentage points in 2021 to +4.3 in 2024. Both women and men increased their use strongly, but men remained ahead. In Tajikistan, the gender gap became much wider. In 2021, men were only 0.9 percentage points behind women, but by 2024 men were ahead by 18.5 percentage points. This was mainly because men's digital payment use increased from 32.6% to 58.8%, while women's use increased only from 33.6% to 40.3%.

Table 2. Gender gaps in digital payments and account ownership in four Central Asian countries, 2021 and 2024

Country	Payments 2021	Payments 2024	Change	Account 2021	Account 2024	Change
Kazakhstan	-8.2	+6.2	+14.3 (±7.9)	-5.3	+2.8	+8.1 (±7.5)
Kyrgyzstan	+2.9	+4.3	+1.4 (±9.7)	+2.8	+2.1	-0.6 (±9.6)
Tajikistan	-0.9	+18.5	+19.5 (±10.6)	+0.2	+15.7	+15.5 (±10.8)
Uzbekistan	+12.6	+2.7	-9.9 (±10.7)	+11.4	-2.2	-13.6 (±10.7)

Note: Payments = made or received a digital payment; Account = account ownership. All values are gaps, men minus women, in percentage points; changes are given with approximate 95% margins of error. Source: Global Findex Database 2025; author's calculations.

The margins of error for the neighbours are wider than for Kazakhstan, because their levels are closer to 50%, where survey estimates are least precise, and their design effects are larger. Two of the four stories are clear in the data. Tajikistan's widening is larger than its margin both for digital payments and for accounts. Uzbekistan's narrowing is clear for accounts (-13.6 ±10.7) and for saving any money (-11.1 ±9.9), but for digital payments the change of -9.9 is just inside its margin of 10.7, so on that indicator alone it could still be survey noise. In Kyrgyzstan none of the changes is close to its margin. Figure 2 shows the digital payment gaps with their margins.

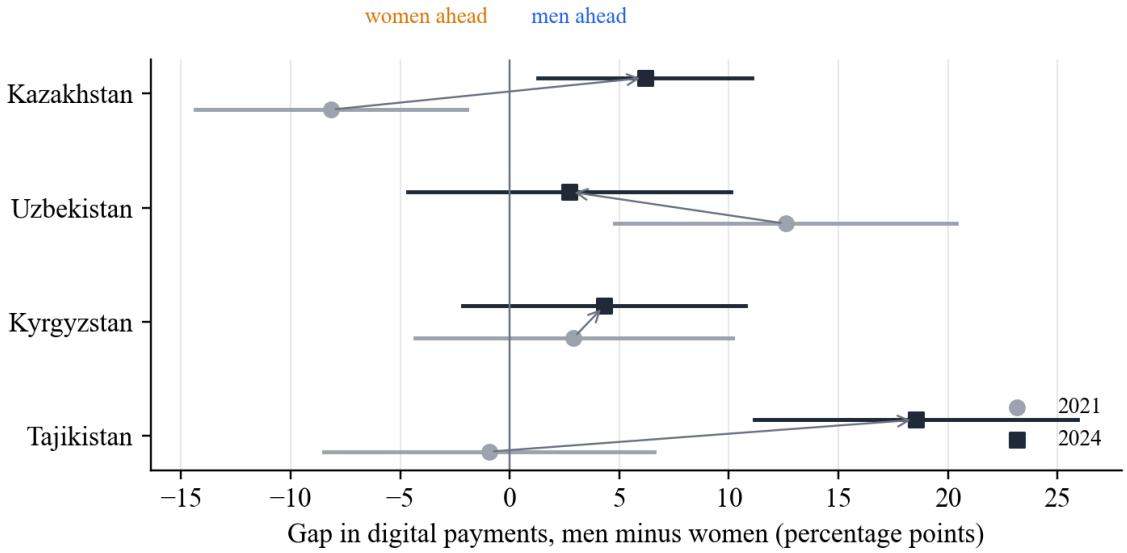


Figure 2. Gender gap in digital payments, men minus women, 2021 and 2024. Lines show approximate 95% margins of error.

What changed inside Kazakhstan

Digital payments can be made or received, and the two parts moved differently (Figure 3). The share of women who made a digital payment changed little, from 70.5% to 68.4%, while the share of men rose from 61.3% to 72.6%. The share who received a digital payment rose for both, by 9.8 points for women and by 24.4 points for men.

Received payments include wages, government transfers and pensions, and money sent by family. Two of these moved very differently for women and men. The share of men who received wages into an account rose from 36.0% to 58.9%, and the share of women from 31.1% to 44.1%. More men than women receive wages in the first place: in 2024, 66.6% of men and 49.5% of women said they had received wages in the past year. At the same time, the share of women who received a government transfer or pension fell from 34.8% to 25.4%, a fall of 9.4 points (±6.6), while among men it fell less, from 23.1% to 19.6%. In 2021 women had been 11.8 points more likely than men to receive such payments.

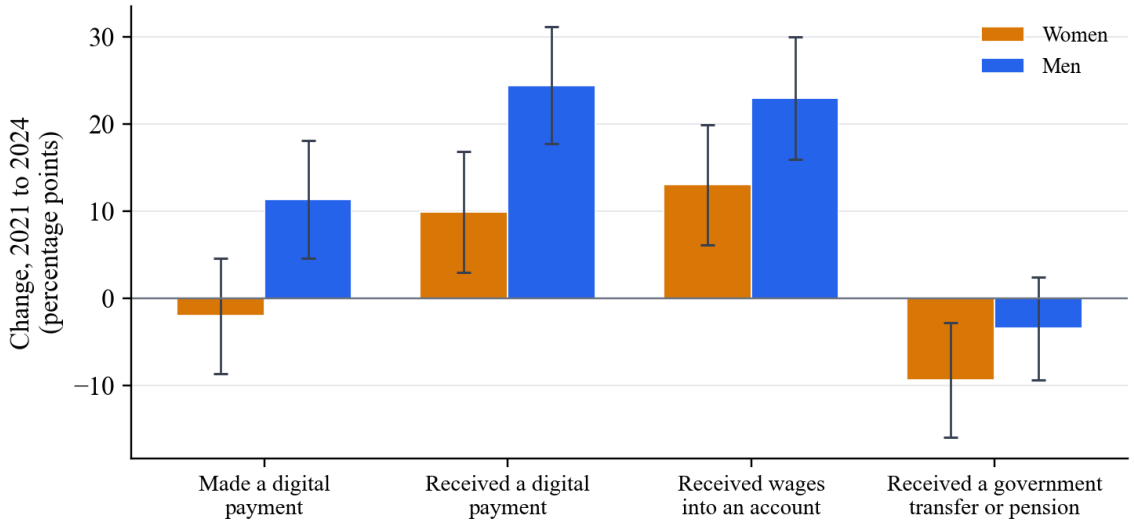


Figure 3. Kazakhstan: change between 2021 and 2024 in four indicators, women and men. Lines show approximate 95% margins of error.

Discussion

The first result is that Kazakhstan's reversal did not happen because women dropped out of digital finance. In 2024, 82.1% of women made or received a digital payment, almost exactly as many as in 2021. The gap changed because men, who were behind, moved past them. This matters for how the change is described. A headline saying that women are now behind would be true, but it would suggest a loss that did not happen.

The second result is about where the change came from. The two indicators that moved most, wages into an account for men and government transfers for women, both concern how people receive money. Men receive wages more often than women, so when more wages are paid into accounts, men's digital payment figures rise more. Women in Kazakhstan were more likely than men to receive government transfers or pensions, and in 2024 fewer women received them than in 2021. The 2021 round was carried out during the COVID-19 pandemic, when, as the Findex report notes, mobility restrictions and health policies increased demand for digital services [7], so part of women's 2021 lead may have depended on payments that were larger or more common at that time. The data cannot show this directly. They only show that both changes happened in the same three years as the reversal.

If this reading is right, the gender gap in Kazakhstan's digital payments may now depend more on the labour market than on access to accounts, because almost everyone already has one. Experimental evidence suggests that the channel through which money arrives matters for women. In India, women whose public sector wages were paid directly into their own bank accounts, together with training in how to use them, worked more and changed their views about women's work [11]. In Kenya, the effects of mobile money on poverty were larger for female-headed households [12]. Both studies are from countries very different from Kazakhstan, but they support the idea that it matters where and to whom payments go.

The neighbours show that there is no single Central Asian pattern. Tajikistan moved the way Kazakhstan did, with men pulling ahead, but from a much lower level of use. Uzbekistan moved the other way: women caught up on accounts, saving and debit cards, and the earlier male advantage almost disappeared. Kyrgyzstan grew fast for both women and men without a change in the gap. Earlier work found that the region as a whole is behind Central and Eastern Europe for women and men alike [6]. My results add that, inside the region, the direction of the

gender gap changed in different ways in just three years. For Kazakhstan, this means the gender gap needs to be tracked in every survey round and for each type of payment, because a national average of 87% account ownership can hide a reversal like this one.

Limitations

This study has several limitations. First, it uses published country averages, not individual answers, so it cannot test whether the women and men who stopped or started using digital payments are the same people who changed how they receive wages or transfers. Second, the margins of error are my own approximations based on an even split between women and men; the official standard errors would be somewhat different. Third, mobile money was not measured in Kazakhstan in 2021, so changes in mobile money accounts could only be described for 2024. Fourth, Uzbekistan's 2024 sample excluded Karakalpakstan. Finally, the comparison covers only two rounds three years apart, and the 2021 round was carried out in an unusual pandemic year, so the next Findex round is needed to see whether the men's lead in Kazakhstan lasts.

Conclusion

Between 2021 and 2024, the gender gap in digital payments in Kazakhstan turned from a lead for women of 8.2 percentage points into a lead for men of 6.2 points. The change is larger than its approximate margin of error, and accounts and saving moved in the same direction. Women's use did not fall; men's use grew by 14.7 points. In the same period, more men started receiving wages into an account, and fewer women received government transfers or pensions. Kazakhstan's neighbours followed three different paths: a widening gap in Tajikistan, a narrowing gap in Uzbekistan and almost no change in Kyrgyzstan. For a country where almost every adult already has an account, the gender gap in digital finance may depend more on how people receive their money than on who has an account.

Data availability

All data are publicly available from the World Bank Global Findex Database 2025 [3]. The table of all values used in this paper and the script that calculates every number and margin of error are available from the author on request.

AI and tools disclosure

I collected the account and digital payment figures for the four countries from the Global Findex database myself and wrote the description of the results. My mentor added the other indicators, checked all values against the World Bank data service, estimated the margins of error and helped prepare the figures and the rest of the text. Generative AI tools were used for language editing and to help with the calculations and figures in Python. No data, results or citations were generated by AI: all numbers come from the World Bank database, and all cited sources were checked against the originals. I have reviewed the final manuscript and take full responsibility for it.

References:

1. Demirgüç-Kunt A., Klapper L., Singer D. Financial inclusion and legal discrimination against women: evidence from developing countries // World Bank Policy Research Working Paper. 2013. No. 6416. DOI: 10.1596/1813-9450-6416.

2. Antonijević M., Ljumović I., Ivanović Đ. Is there a gender gap in financial inclusion across countries? // Journal of Women's Entrepreneurship and Education. 2022. No. 1-2. P. 79-96. DOI: 10.28934/jwee22.12.pp79-96.

3. World Bank. Global Findex Database 2025 [Electronic resource]. URL: https://databank.worldbank.org/source/global-financial-inclusion (accessed: 06.10.2026).

4. Bekturganova M., Kurmasheva M. Structural barriers and opportunities for financial inclusion of women in Kazakhstan // Eurasian Journal of Economic and Business Studies. 2025. Vol. 69, No. 2. P. 19-33. DOI: 10.47703/ejebs.v69i2.492.

5. Kireyeva A.A., Kurmasheva M.T. Assessment of the level of financial inclusion of women in Kazakhstan: an empirical analysis of structural factors // Economy: Strategy and Practice. 2025. Vol. 20, No. 2. P. 104-121. DOI: 10.51176/1997-9967-2025-2-104-121.

6. Poghosyan T. The anatomy of the financial inclusion gap in the Caucasus and Central Asia // IMF Working Paper. 2023. No. 2023/109. DOI: 10.5089/9798400243431.001.

7. Demirgüç-Kunt A., Klapper L., Singer D., Ansar S. The Global Findex Database 2021: Financial inclusion, digital payments, and resilience in the age of COVID-19. Washington, DC: World Bank, 2022. DOI: 10.1596/978-1-4648-1897-4.

8. Klapper L., Singer D., Starita L., Norris A. The Global Findex Database 2025: Connectivity and financial inclusion in the digital economy. Washington, DC: World Bank, 2025. DOI: 10.1596/978-1-4648-2204-9.

9. World Bank. The Global Findex Database 2021: Survey methodology [Electronic resource]. URL: https://thedocs.worldbank.org/en/doc/fe026d04da1eb78f5ce8bb5e1529d60b-0430062023/original/Findex2021-22-Methodology.pdf (accessed: 06.10.2026).

10. World Bank. The Global Findex Database 2025: Survey methodology [Electronic resource]. URL: https://thedocs.worldbank.org/en/doc/be6615202d1f08a25855c8ac2d615122-0050012025/related/Survey-methodology.pdf (accessed: 06.10.2026).

11. Field E., Pande R., Rigol N., Schaner S., Troyer Moore C. On her own account: how strengthening women's financial control impacts labor supply and gender norms // American Economic Review. 2021. Vol. 111, No. 7. P. 2342-2375. DOI: 10.1257/aer.20200705.

12. Suri T., Jack W. The long-run poverty and gender impacts of mobile money // Science. 2016. Vol. 354, No. 6317. P. 1288-1292. DOI: 10.1126/science.aah5309.

UDC 336.22:334.722(574)

Amangeldi Aidana
Grade 11 student
Supervisor: Onayev Ainabek
Republican Physics and Mathematics School
(Astana, Kazakhstan)

DID SMALL BUSINESS IN KAZAKHSTAN REALLY SHRINK IN 2026? THE NEW TAX CODE, REGIONAL TAX RATES AND THE FALL IN THE NUMBER OF INDIVIDUAL ENTREPRENEURS

Abstract: From 1 January 2026 Kazakhstan has a new Tax Code. The basic rate of the simplified declaration regime for small business became 4%, and every local council (maslikhat) could lower or raise it by up to half, so different regions chose different rates. At the same time, official statistics showed that the number of active small and medium enterprises (SMEs) fell by 8.8% between 1 January and 1 September 2026. In this study I asked whether the regions that chose lower rates kept more of their businesses, and where the fall actually came from. I used 28 monthly releases of the Bureau of National Statistics (June 2024 - September 2026) and the list of maslikhat decisions for all 20 regions. In the same months of 2025 the number of active SMEs grew by 5.0%, so the change in 2026 was about 13.7 percentage points worse than a year before. The whole fall was in individual entrepreneurs, whose number dropped by 11.3% (203,398 units), while legal entities grew by 3.2%. Almost all of the fall happened between January and May 2026 and then stopped. Three groups of activities, transport (taxi and couriers), renting of housing and personal services, held 30% of individual entrepreneurs but gave 74% of the fall, and these are the activities allowed in the new special tax regime for the self-employed, which can be used only by people who are not registered as entrepreneurs. The regional rates did not explain the fall: regions with a 2% and a 3% rate in their regional centre lost a similar share of SMEs (median -7.8% and -9.2%, $p = 0.34$). My conclusion is that most of the fall is a change in how people are registered after the new tax rules, and not a mass closure of businesses, and that SME counts for 2026 cannot be compared with earlier years without taking this into account.

Keywords: small and medium enterprises; individual entrepreneurs; self-employment; tax reform; local taxes; business register; Kazakhstan

1. Introduction

I have always been interested in how economic decisions are made in Kazakhstan and whether they actually change something. The new Tax Code, which came into force on 1 January 2026, gave me a rare example where such a decision was made separately in each region. For small businesses that use the simplified declaration, the Code sets a basic rate of 4% of income, but local representative bodies (maslikhats) can lower or raise it by up to 50% depending on the type of activity and the location [1]. Most of them lowered it. For example, the maslikhats of Astana and Almaty set 3%, and Shymkent set 2% [2]. So every region made its own choice, and it is possible to compare what happened to business in regions with different rates.

When my mentor and I looked at the official statistics, the first picture was alarming. According to the Bureau of National Statistics, the number of active SMEs in Kazakhstan was

2,175,605 on 1 January 2026 and 1,984,663 on 1 September 2026, which is 8.8% less [3]. In some regions the fall was bigger: -13.1% in Kostanay region and -12.3% in West Kazakhstan. If this were a real closure of almost 200 thousand businesses in eight months, it would be one of the most important economic events of the year.

There are studies about how simpler and lower taxes change the number of registered small firms. In Brazil, the SIMPLES scheme, which reduced and simplified taxes for micro-firms, increased the share of firms that operate formally [4, 5]. A review of reforms in developing countries found that making registration easier usually gives only a modest increase in the number of formal firms [6]. Tax thresholds and special regimes also decide which firms register and in what form [7]. From these studies I expected that a lower regional rate would keep more businesses registered.

However, the statistics releases themselves contain a footnote. It says that the fall in the number of individual entrepreneurs is connected with their transfer to the special tax regime for the self-employed, introduced by the new Tax Code from 1 January 2026 [3]. This changed my question. In this study I asked:

• How did the number of active SMEs, individual entrepreneurs and legal entities change between January and September 2026, compared with the same months of 2025?

• Did the regions with lower simplified declaration rates lose fewer SMEs?

• In which types of activity did individual entrepreneurs disappear, and does this match the activities allowed for the self-employed?

2. What changed in the tax rules

Three changes in the new Tax Code are important for this study.

The simplified declaration. Individual entrepreneurs and small companies with income up to 600,000 monthly calculation indices (MCI) per year can use the special tax regime based on the simplified declaration [8]. In the new Code the rate is 4%, and a maslikhat can change it by up to 50%, so the possible range is from 2% to 6%. The decision had to be made not later than 1 December 2025 and came into force on 1 January 2026 [1]. The old Tax Code had a different rate and a different set of taxes for this regime, so in this study I compare regions only with each other and not with the old rate.

The regime for the self-employed. The new Code created a separate special tax regime for the self-employed. A self-employed person is defined as an individual who is not an individual entrepreneur, does not employ workers, earns not more than 300 MCI per month and works in one of the activities from a government list [9]. The individual income tax rate in this regime is 0%, and the person pays only social payments through a special mobile application [10]. The list approved by the Government contains 40 activities, including taxi services, courier activity, delivery of ready food, renting of own or rented housing, hairdressers and beauty salons, and many kinds of repair [11]. The important point is that a person cannot be self-employed and an individual entrepreneur at the same time.

The end of the patent and mobile application regimes. The old regimes based on a patent and on a special mobile application were cancelled. Entrepreneurs who used them had to submit a notification about their new tax regime by 1 March 2026. Those who did not were removed from registration as individual entrepreneurs automatically, with the date of removal 1 January 2026 [12].

3. Data and methods

Statistics. I used the monthly releases of the Bureau of National Statistics "Main indicators of the number of entities in the Republic of Kazakhstan" [3]. From each release I took

Table 1 (registered and active legal entities, individual entrepreneurs and SMEs for the whole country), Table 5 (individual entrepreneurs by type of activity) and Table 9 (registered and active SMEs by region). I downloaded all 28 releases from 1 June 2024 to 1 September 2026 with Python. In the Bureau's register, a unit is "active" if its situation code shows economic activity, and these codes are updated every month, partly with information from the State Revenue Committee about inactive taxpayers [13]. In all calculations I used active units.

Regional rates. For each of the 17 regions and three cities of republican significance I took the rate set by the maslikhat of the regional centre (for example, Semey for Abai region and Taldykorgan for Zhetisu) from a list of maslikhat decisions updated on 7 August 2026 [2], and checked that every decision, its date and number were in this list. Ten regions had 2% in their regional centre and ten had 3%. Because district maslikhats inside a region could choose a different rate, I also calculated the share of the listed maslikhats of each region that chose 2%.

Comparison with the previous year. The number of entrepreneurs changes during the year, so a fall from January to September could partly be seasonal. To avoid this, I compared the change from 1 January to 1 September 2026 with the change over the same months of 2025. I call the difference between the two changes the "excess change" of a region. To compare regions with 2% and 3%, I used the Mann-Whitney test, and for the share of 2% maslikhats and for the size of the region I used the Spearman correlation. With only 20 regions these tests can find only quite strong relationships.

Activities. Table 5 does not give individual entrepreneurs by region, only by section of the classification of economic activities. In three sections the main activities of individual entrepreneurs are on the list for the self-employed: transport and storage (taxi and courier activity), real estate (renting of housing) and other personal services (hairdressers, repair of personal and household goods) [11]. I call them the "self-employed sections" and compare them with all other sections. These sections also contain activities that are not on the list, such as freight transport, so this grouping is only approximate.

4. Results

4.1. The fall was only in individual entrepreneurs

Table 1 shows the main result. Between 1 January and 1 September 2025 the number of active SMEs grew by 5.0%, but over the same months of 2026 it fell by 8.8%. The whole fall was in individual entrepreneurs, whose number decreased by 203,398, which is even a bit more than the fall in all SMEs (190,942). The number of active legal entities grew in 2026 by 3.2%, faster than in 2025.

Table 1. Active entities in Kazakhstan, January-September 2025 and 2026

Active units	1 Jan 2025	1 Sep 2025	Change 2025, %	1 Jan 2026	1 Sep 2026	Change 2026, %
Small and medium enterprises	2,071,657	2,174,309	+5.0	2,175,605	1,984,663	-8.8
Individual entrepreneurs	1,705,517	1,805,526	+5.9	1,799,836	1,596,438	-11.3
Legal entities	426,096	428,987	+0.7	436,864	450,688	+3.2

Source: Bureau of National Statistics, Tables 1 and 9 of the monthly releases. Legal entities include large ones, so they are not a part of SMEs only.

The monthly data show when it happened (Figure 1). Until December 2025 the number of individual entrepreneurs was growing. In 2026 it fell every month from January to May: by 29,611 in January, 37,395 in February, 43,089 in March and 92,574 in April (between the releases for 1 April and 1 May). After that it stayed almost the same, around 1.59-1.60 million. The largest monthly fall appears in the release for 1 May, two months after the deadline. I think this is the time the statistical register needed to receive the tax authorities' decisions, but I could not confirm it.

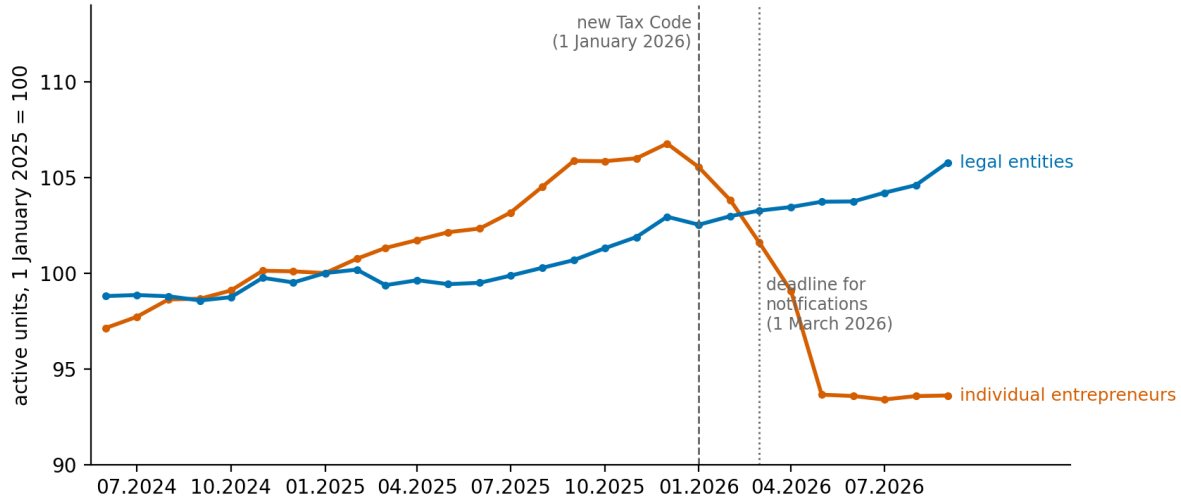


Figure 1. Active individual entrepreneurs and legal entities in Kazakhstan, June 2024 - September 2026 (1 January 2025 = 100)

4.2. The regional rates did not explain the fall

All 20 regions lost SMEs in 2026, from -5.1% in Pavlodar region to -13.1% in Kostanay region. In the same months of 2025 all of them except Karaganda region had grown. Table 2 shows that the regions with a 2% rate in their regional centre lost a median of 7.8% of their SMEs, and the regions with 3% lost 9.2%. This difference goes in the expected direction, but it is small compared with the differences inside each group, and it is not statistically significant (Mann-Whitney p = 0.34). The result was the same for the excess change, which takes into account how fast each region was growing in 2025 (median -11.0 and -14.4 percentage points, p = 0.47), and for the share of maslikhats that chose 2% (Spearman ρ = 0.18, p = 0.44). It is also important that before the reform the two groups grew at the same speed (median +3.6% and +3.7% in 2025, p = 0.68).

Table 2. Change in active SMEs (1 January - 1 September) by the regional rate

Group of regions	Number of regions	Change 2025, % (median)	Change 2026, % (median)	Excess change, p.p. (median)
2% in the regional centre	10	+3.6	-7.8	-11.0
3% in the regional centre	10	+3.7	-9.2	-14.4
Mann-Whitney p		0.68	0.34	0.47

Rates are for the maslikhat of the regional centre, or of the city for Astana, Almaty and Shymkent. Excess change is the change in 2026 minus the change in 2025.

The only regional feature that was related to the excess change was the size of the region (Figure 2). Larger regions lost more compared with their growth in 2025 (Spearman ρ = -0.46, p = 0.04). The excess change was -17.2 percentage points in Almaty, -17.0 in Astana and -19.4 in Almaty region, while in Pavlodar region it was -6.8 and in Karaganda region -7.3. Astana and Almaty both had a 3% rate, so part of the small difference between the 2% and 3% groups may come from the big cities being in the 3% group.

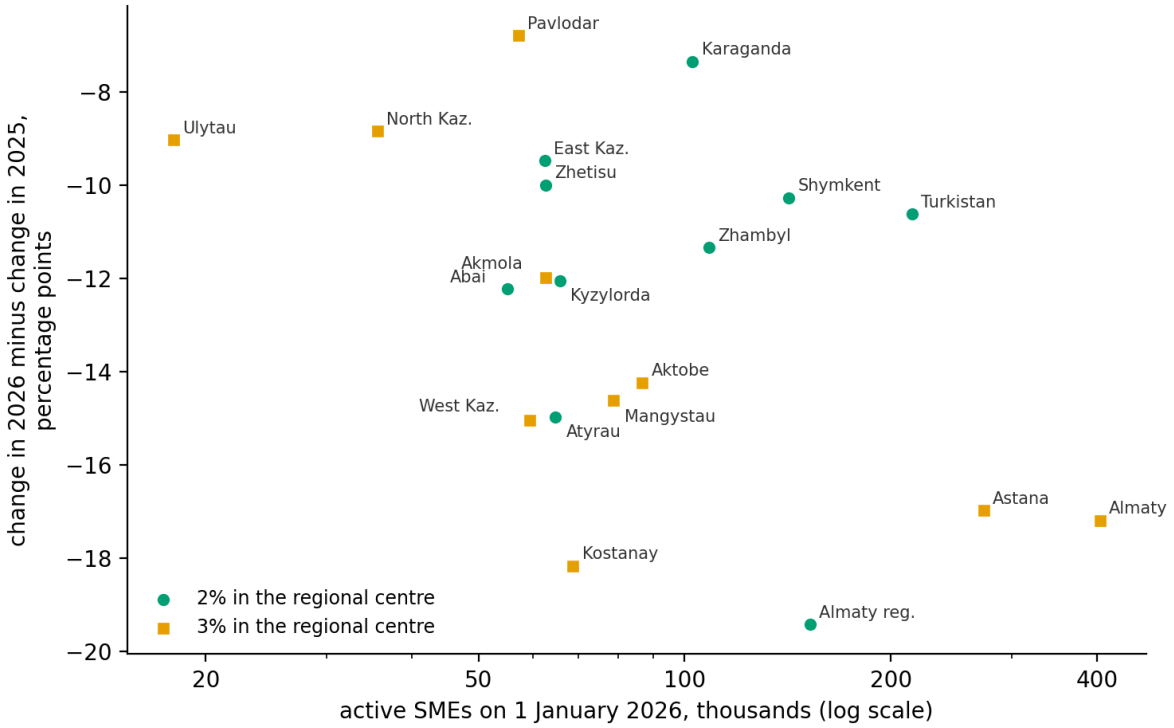


Figure 2. Excess change in active SMEs (change in 2026 minus change in 2025) against the number of active SMEs in the region

4.3. Individual entrepreneurs disappeared in the activities allowed for the self-employed

Figure 3 shows the change in active individual entrepreneurs by type of activity. The biggest fall was in transport and storage, where the number of individual entrepreneurs fell from 173,432 to 90,844, by 47.6%. This is especially interesting because in the same months of 2025 the same section had grown by 48.1%. Real estate fell by 20.1% and other personal services by 17.4%, while in 2025 they grew by 6.6% and 0.7%. Together these three self-employed sections held 30% of active individual entrepreneurs on 1 January 2026, but they gave 74% of the fall (149,590 of 203,398). In all other sections together the number of individual entrepreneurs fell by only 4.3%. Information and communication, which has no activities on the list, grew by 8.2% in 2026, almost the same as in 2025 (9.2%).

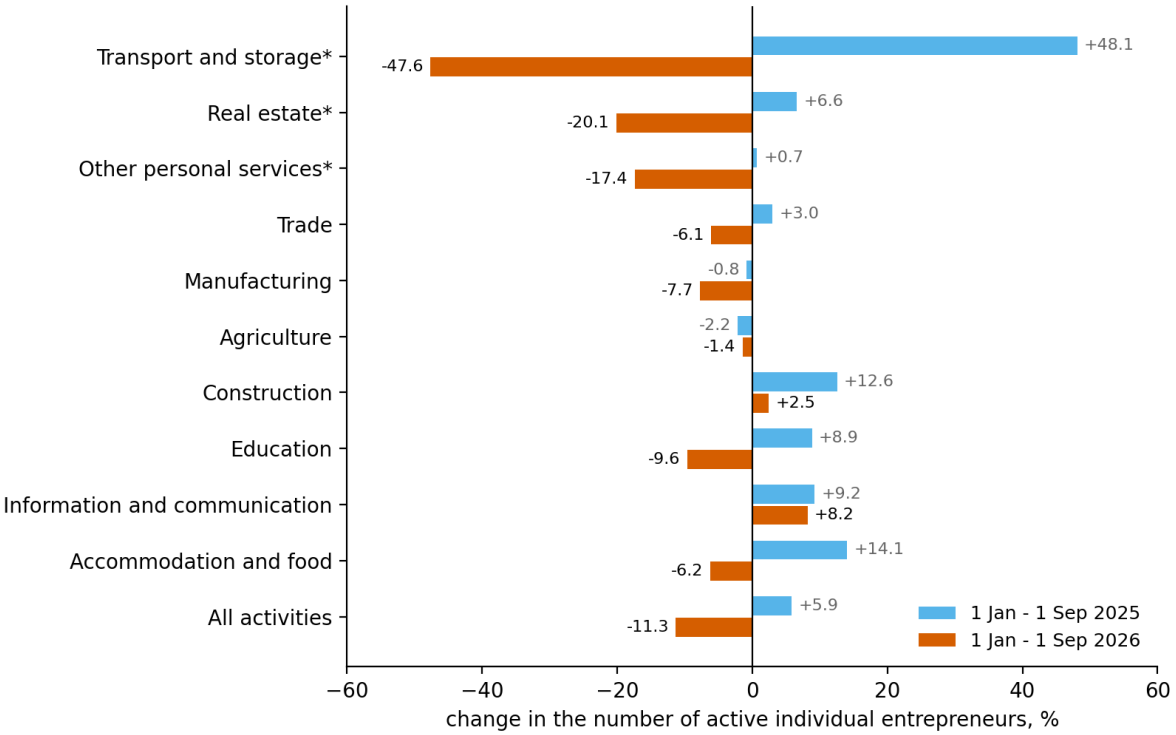


Figure 3. Change in active individual entrepreneurs from 1 January to 1 September, by type of activity, 2025 and 2026 (* sections where the main activities of entrepreneurs are on the list for the self-employed)

Trade also lost 39,507 entrepreneurs (-6.1%), which is a large number but a small share of this very big section. Retail trade through network marketing is the only trade activity on the list [11], so trade entrepreneurs may have left for other reasons, for example because the patent regime was cancelled.

4.4. How many people became self-employed

The Bureau's statistics do not include the self-employed, because they are not entrepreneurs. According to the State Revenue Committee, by the beginning of March 2026 about 346 thousand people had chosen the regime for the self-employed, and 265 thousand of them registered their activity for the first time [14]. The remaining 81 thousand had been registered before in some form, most likely as entrepreneurs. These numbers are from a news report, and I could not check them in an official statistical table. Still, they show that the number of people working in the new regime was larger than the fall in the number of individual entrepreneurs.

5. Discussion

At the beginning I expected to find that regions with lower rates kept more of their businesses. The data did not show this. The fall in SMEs in 2026 was almost entirely a fall in individual entrepreneurs, it happened in a few months around the deadline of 1 March, and it was concentrated in the activities that the new regime for the self-employed allows. The Bureau of National Statistics gives the same explanation in its footnote [3]. Together this suggests that most of the "lost" entrepreneurs did not close their business. They changed the way they are registered: a taxi driver or a person renting out a flat can now work as self-employed with 0% income tax and no registration as an entrepreneur [9, 10]. Some of them were also removed automatically because they did not submit a notification in time [12]. I cannot say how many of

the 203 thousand continued working and how many really stopped, because the data do not follow individual people.

This also explains why the regional rates did not matter. A difference of one percentage point in the simplified declaration rate is small compared with the difference between paying 2-3% as an entrepreneur and paying 0% income tax as a self-employed person. For people who could move to the new regime, the regional rate was probably not the main factor. For the businesses that stayed entrepreneurs, the effect of the rate may exist, but it is hidden by the much larger change in registration. The larger fall in big cities fits this explanation, because taxi, delivery and renting of housing are probably more common in Astana and Almaty, but I did not have regional data by activity to check it.

The fast growth of transport entrepreneurs in 2025 (+48%) and their fall in 2026 (-48%) look almost like a mirror image. I could not find an official explanation for the growth in 2025, so I only note it. If many of these entrepreneurs were taxi drivers who registered in 2025 and then moved to the self-employed regime in 2026, it would mean that the number of entrepreneurs in this section reflects the tax rules more than the real number of drivers.

From the point of view of someone who uses statistics, the main lesson is that the number of SMEs in 2026 is not comparable with earlier years. If someone reads the -8.8% as a crisis of small business, they will draw the wrong conclusion. When the definition of who is counted changes, the data need to be read together with the new category. In this case the self-employed should be counted next to the entrepreneurs.

Limitations. The data count registered units, not their income or the number of people working. Individual entrepreneurs by activity are available only for the whole country, not by region. I used the rate of the regional centre, although district maslikhats in the same region sometimes chose a different rate and the rate could also depend on the type of activity. There are only 20 regions, so a small effect of the rate could not be found even if it exists. The comparison covers only one year before the reform, and other events in 2026 could also affect the number of entrepreneurs. The number of self-employed is taken from a news report quoting the State Revenue Committee and is not an official statistical series. Finally, the timing of changes in the statistical register may lag behind the real decisions of the tax authorities.

6. Conclusion

The official statistics show an 8.8% fall in the number of active SMEs in Kazakhstan between January and September 2026, while a year earlier the number was growing. I found that the whole fall was in individual entrepreneurs, that it happened between January and May 2026, and that three quarters of it was in the activities allowed for the new regime for the self-employed. The regional simplified declaration rates chosen by maslikhats did not explain the differences between regions. My conclusion is that the fall mainly reflects a change in how people register their activity after the new Tax Code, and not a mass closure of small businesses. To find out how many of these people really continued working, the next step would be to compare the number of self-employed and entrepreneurs by region and activity when such data become available, and to use income and employment data instead of the number of registered units.

Author contribution and use of AI

The author chose the general direction of the study, how economic decisions are made in Kazakhstan, and took part in choosing the research question together with her mentor. The author's mentor collected and checked the data, carried out the analysis, made the figures and helped write and edit the English text. The author read the final text and agreed with its

conclusions. Generative AI tools were used to help with editing and translation. Every number in the paper was calculated from the official sources by the analysis code and checked by the author's mentor.

References:

1. Налоговый кодекс Республики Казахстан от 18 июля 2025 года № 214-VIII. Статья 726. Исчисление налогов по специальному налоговому режиму на основе упрощенной декларации. URL: https://adilet.zan.kz/rus/docs/K2500000214 (accessed: 03.10.2026).

2. PRO1C. Актуальный список маслихатов, понизивших ставку по СНР на основе упрощенной декларации (updated 07.08.2026). URL: https://pro1c.kz/articles/prochee/aktualnyy-spisok-maslikhatov-ponizivshikh-stavku-po-snr-na-osnove-uproshchennoy-deklaratsii/ (accessed: 03.10.2026).

3. Бюро национальной статистики Агентства по стратегическому планированию и реформам Республики Казахстан. Основные показатели количества субъектов в Республике Казахстан: monthly releases on 1 June 2024 - 1 September 2026, Tables 1, 5 and 9. URL: https://stat.gov.kz/ru/industries/business-statistics/stat-org/publications/ (accessed: 03.10.2026).

4. Fajnzylber P., Maloney W.F., Montes-Rojas G.V. Does formality improve micro-firm performance? Evidence from the Brazilian SIMPLES program // Journal of Development Economics. 2011. Vol. 94, No. 2. P. 262-276. DOI: 10.1016/j.jdeveco.2010.01.009.

5. Monteiro J.C.M., Assunção J.J. Coming out of the shadows? Estimating the impact of bureaucracy simplification and tax cut on formality in Brazilian microenterprises // Journal of Development Economics. 2012. Vol. 99, No. 1. P. 105-115. DOI: 10.1016/j.jdeveco.2011.10.002.

6. Bruhn M., McKenzie D. Entry regulation and the formalization of microenterprises in developing countries // The World Bank Research Observer. 2014. Vol. 29, No. 2. P. 186-201. DOI: 10.1093/wbro/lku002.

7. Kanbur R., Keen M. Thresholds, informality, and partitions of compliance // International Tax and Public Finance. 2014. Vol. 21, No. 4. P. 536-559. DOI: 10.1007/s10797-014-9314-3.

8. Налоговый кодекс Республики Казахстан от 18 июля 2025 года № 214-VIII. Статья 723. Условия применения специального налогового режима на основе упрощенной декларации. URL: https://adilet.zan.kz/rus/docs/K2500000214 (accessed: 03.10.2026).

9. Налоговый кодекс Республики Казахстан от 18 июля 2025 года № 214-VIII. Статья 718. Общие положения (специальный налоговый режим для самозанятых). URL: https://adilet.zan.kz/rus/docs/K2500000214 (accessed: 03.10.2026).

10. Налоговый кодекс Республики Казахстан от 18 июля 2025 года № 214-VIII. Статья 720. Порядок исчисления и уплаты индивидуального подоходного налога и социальных платежей в специальном налоговом режиме. URL: https://adilet.zan.kz/rus/docs/K2500000214 (accessed: 03.10.2026).

11. Постановление Правительства Республики Казахстан от 21 ноября 2025 года № 994 «Об утверждении перечня видов деятельности, по которым разрешено применение

специального налогового режима для самозанятых». URL: https://adilet.zan.kz/rus/docs/P2500000994 (accessed: 03.10.2026).

12. Uchet.kz. Не подали уведомление о налоговом режиме до 1 марта 2026 года: что будет дальше? 11 February 2026. URL: https://uchet.kz/news/ne-podali-uvedomlenie-o-nalogovom-rezhime-do-1-marta-2026-goda-chto-budet-dalshe/ (accessed: 03.10.2026).

13. Бюро национальной статистики. Методика по ведению и актуализации информационной системы Статистический бизнес-регистр (attached to the releases). URL: https://stat.gov.kz/upload/iblock/0b6/94cfaiukxjvfu279wf7qw680s56oj95v.docx (accessed: 03.10.2026).

14. NUR.KZ. Переход на новые налоговые режимы завершен: сколько самозанятых в Казахстане. 4 March 2026. URL: https://www.nur.kz/nurfin/fines-and-taxes/2349999-perehod-na-novye-nalogovye-rezhimy-zavershen-skolko-samozanyatyh-v-kazahstane/ (accessed: 03.10.2026).

UDC 338.45:620.91(574)

Cheng Che
DBA Student
Al-Farabi Kazakh National University
(Almaty, Kazakhstan)

BUILD, BUY, OR PARTNER? ENTRY MODE SELECTION AND RISK GOVERNANCE FOR WIND POWER INVESTMENT IN KAZAKHSTAN

Abstract. Foreign wind power investors in Kazakhstan can build, acquire during construction, or develop with partners. Each route relocates uncertainty. This article develops a comparative qualitative framework for choosing among these modes and governing their risks. Public records concerning Zhanatas, Shokpar, Dostyk, two different Arkalyk assets, and the announced 1 GW program form the evidence base. The RIGOR framework assesses Rights completeness, Information reliability, Governance complementarity, Option value, and Residual risk. An integrated acquisition model then links eight diligence domains to legality, loss pathway, liability allocation, and post-closing leadership. Public facts, allegations, and practitioner interpretations are separated. Greenfield development suits investors with patient development capability and a need for design control. Construction-stage acquisition can save time when rights are formed and evidence is auditable, but it concentrates interface and completion risk. Joint development suits high local complementarity and program optionality when governance rights are explicit. Dostyk also shows how publicly reported credit insurance and RMB project finance can support an acquired project without replacing asset diligence. The article avoids confidential transaction data.

Keywords: entry mode; wind power; Kazakhstan; greenfield development; construction-stage acquisition; joint development; due diligence; project finance; risk governance

1. Introduction

Renewable-market entry is a choice about which uncertainties an investor will own. A greenfield investor accepts development uncertainty for design control. An acquirer pays for elapsed time but inherits earlier decisions. A joint developer shares resources but must govern shared control. The relevant question is which mode creates the most governable combination of rights, evidence, capability, and residual risk.

Transaction cost economics links internalization to contractual hazards and asset specificity [1], while entry-mode research connects those hazards to control [2]. Greenfield and acquisition choices also reflect capability transfer, target availability, integration cost, and institutional distance [3] [4]. Real-options research explains why staged commitment can preserve expansion value under uncertainty [5] [6] [7].

In an emerging market, the target may be partly constructed rather than operational. This brownfield form combines local embeddedness with unfinished work [8]. Its success depends on entry choice, integration design, and the conversion of deal completion into operating value [9] [10] [11].

Wind projects add high capital intensity, site-specific resources, grid dependence, and policy exposure [12] [13]. Capital providers accept different combinations of construction and revenue risk [14]. Entry and financing must therefore be considered together.

Kazakhstan offers observable examples of greenfield development, construction-stage acquisition, and program partnership. This article asks:

1. How should a foreign wind power investor select among building, buying, and partnering in Kazakhstan? 2. How should due diligence and post-entry governance be integrated when the selected route is a construction-stage acquisition?

The comparison uses public documents. It excludes transaction prices, sellers, acquisition dates, debt terms, and internal information. Experience-based inferences are labelled practitioner interpretation so that fact, allegation, and judgment remain separate.

2. Institutional and analytical foundations

2.1 The project as a bundle of rights and interfaces

A wind project is a bundle of rights, assumptions, contracts, relationships, works, claims, and obligations. These mature at different speeds. Land may be valid while the grid solution is provisional. Equipment may be contracted while substation responsibility is unclear.

Kazakhstan's renewable law defines support mechanisms [15]. Auction materials published by the Financial Settlement Centre explain how tariff rights are formed [16]. These rules reduce uncertainty, but site, grid, corporate history, and execution remain project-specific.

Maturity should not be reduced to construction percentage. Rights completeness asks whether reliable evidence proves the ability to build, connect, sell, collect cash, operate, and enforce contracts. Physical progress and rights completeness may diverge.

2.2 Three entry modes

Greenfield development starts before core rights and design are assembled. It provides design control but carries delay and failure risk.

Construction-stage acquisition follows material development decisions but precedes stable operation. It saves time while transferring frozen design choices, work quality, contractor interfaces, permit history, and stakeholder commitments.

Joint development combines complementary development, capital, technology, public coordination, or grid capability. It can preserve options but may create deadlock and blurred accountability.

A project may move across modes. Entry is therefore a governed sequence of commitments.

2.3 Comparative qualitative design and evidence protocol

The comparative case design selects projects with different combinations of mode, maturity, partnership, and finance. Country context follows the International Energy Agency review [17]. Project evidence comes primarily from government records, multilateral lenders, institutional announcements, and insurer reporting.

Three evidence classes are used throughout:

Public fact is traceable to an official record or attributable institutional disclosure.

Public allegation records an attributed complaint without establishing its merits.

Practitioner interpretation is the author's experience-based analysis, not confidential evidence or verified project fact.

This protocol prevents observed outcomes from becoming unsupported causal claims.

The case selection reflects the author's professional experience with greenfield development, construction-stage acquisition, and portfolio development in Kazakhstan. That experience informs the research questions and interface analysis, but it is treated as positionality rather than independent proof of a project fact. No confidential transaction record or unpublished performance result is used.

Table 1. Comparative case design and evidence boundary

Case or program	Publicly observable mode	Analytical purpose	Evidence boundary
Zhanatas, 100 MW	Jointly owned greenfield development	Control, partner complementarity, and multilateral bankability	Public lender and project disclosures only
Shokpar, 100 MW	Adjacent co-development	Repeat-project learning and the need to reset site-specific diligence	Public lender disclosures and an attributed public complaint
Dostyk, 50 MW	Acquisition during construction	Completion interfaces, acquisition diligence, insurance, and RMB finance	Public insurer and project reporting, with commercial deal details excluded
Arkalyk-Kostanay, 48 MW as built	Auction-originated greenfield project	Separation of tender scope, completed asset, and ownership-chain evidence	Official auction and regional records, without inferring later ownership from name alone
Arkalyk-Akmola, 15 MW	Subproject within the CPIH 220 MW portfolio	Identity control and portfolio record matching	AIIB portfolio and non-technical summary only
Announced 1 GW program	Program-level partnership	Governance of scale, storage, infrastructure, and phased commitment	Dated public announcements, not evidence of final investment or completion

3. The RIGOR entry mode selection framework

3.1 Five decision dimensions

The proposed RIGOR framework evaluates five dimensions before an investor chooses to build, buy, or partner.

R: Rights completeness. Land, permits, tariff position, grid rights, approvals, and contracts must be valid, transferable, consistent, and sufficient for the next gate. Low completeness favors greenfield work if the investor can create the missing rights. Acquisition requires rights that survive control transfer.

I: Information reliability. A large data room may still fail to prove the design basis. Reliable evidence has origin, version control, cross-discipline reconciliation, and independent testing where material. Weak evidence favors staged commitment or conditions.

G: Governance complementarity. A partner matters when needed capability, such as public coordination, permitting, grid knowledge, finance, or localization, is hard to contract. Complementarity needs clear decision rights.

O: Option value. Waiting, expansion, replication, refinancing, and exit can have value. A first project can improve a second, while shared infrastructure can support adjacent capacity. Options are strongest when milestones reveal information.

R: Residual risk. Remaining shareholder exposure should be stated as delay, lost energy, added capital, reduced revenue, liability, reputational harm, or loss of control. Paper allocation fails if the responsible party lacks capacity or enforcement is weak.

3.2 From dimensions to mode choice

RIGOR is a structured comparison, not a mechanical score. Weak early rights may repel a financial investor but create value for a developer. High complementarity supports partnership only when reserved matters, reporting, funding, and deadlock rules are credible.

Table 2. RIGOR entry mode selection framework

RIGOR condition	Build	Buy during construction	Partner
Rights completeness	Low at entry, but development pathway is controllable	Substantially formed and transferable	Mixed or program-level rights that partners can assemble jointly
Information reliability	Investor creates the evidence base	Existing evidence must be auditable and reconciled	Partners disclose and validate evidence through shared gates
Governance complementarity	External inputs can be contracted	Seller legacy can be separated from future control	Complementary capability is material and continuing
Option value	Value comes from design freedom and future replication	Value comes from saved time and visible progress	Value comes from staged scale, shared infrastructure, and expansion rights
Residual risk	Development delay and failure	Completion, interface, inherited compliance, and integration	Deadlock, diluted control, funding asymmetry, and partner performance
Preferred control mechanism	Internal development gates	Conditions precedent, warranties, remedies, and integration plan	Shareholders agreement, reserved matters, information rights, and staged funding

Build when development capability creates advantage. Buy when rights and evidence are mature and saved time exceeds integration cost. Partner when complementarity and optionality are high, but formalize governance before major commitment.

4. Integrated due diligence for construction-stage acquisition

4.1 Why conventional workstreams are insufficient

Specialist diligence can conceal risks between workstreams. A grid delay can alter sequencing, trigger claims, increase funding, and breach a financing condition. Every adviser may be locally correct while the decision remains incomplete.

The original 4L-8D model applies four lenses: **legality**, whether a right is valid; **loss pathway**, how a defect affects schedule, cash, energy, or liability; **liability allocation**, who bears an enforceable consequence; and **post-closing leadership**, who manages it after transfer.

Across eight domains, the output is one evidence map, red-flag register, adjusted base case, conditions schedule, and integration plan.

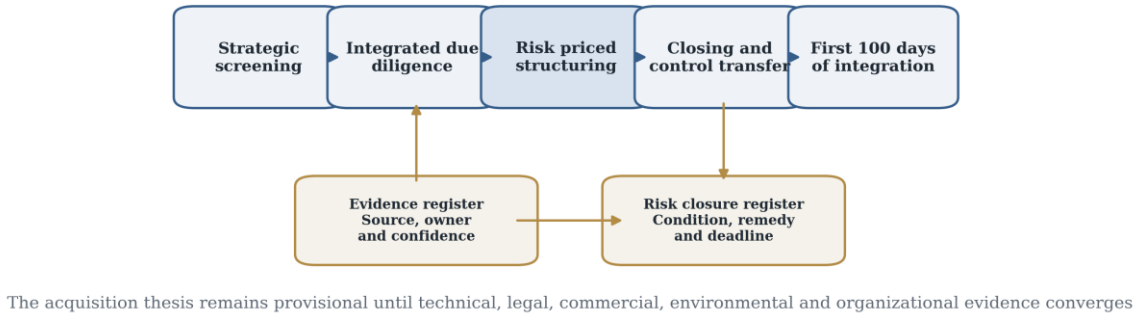


Figure 1. Evidence based acquisition and integration cycle

Source: Developed by the author.

4.2 The evidence based acquisition cycle

The cycle starts by reconciling legal entity, ownership, name, coordinates, capacity, tariff record, grid point, and key contracts. This identity sheet prevents records from one asset being assigned to another, especially where transliteration and portfolio announcements create similar labels.

Next, every assumption is linked to a source, owner, validity, and verification status. Red flags are issues that can defeat the thesis, finance, closing, or selected mode. Findings are translated into schedule and cash effects, then allocated through correction, contract, insurance, conditions, contingency, or acceptance.

The cycle continues after closing. Document control, delegations, contractor instructions, permits, reporting, safety, stakeholder duties, and lender interfaces must transfer without stopping the works. Planning starts before approval. Operating evidence then improves future screening.

Table 3. Integrated acquisition due diligence model

Domain	Core evidence question	Principal loss pathway	Typical governance response
Corporate and control	Does the target own the represented project and can control lawfully transfer?	Invalid title, undisclosed obligation, or blocked control	Registry verification, authority matrix, conditions, and control-transition plan
Land and permits	Are site rights and approvals valid for the built design and full operating life?	Stop-work order, redesign, access loss, or remediation	Legal opinion, cadastral reconciliation, permit tracker, and renewal owner
Resource and energy yield	Are measurements, losses, wake effects, and availability assumptions independently supportable?	Revenue shortfall and covenant pressure	Independent engineer review, sensitivity cases, and performance monitoring
Grid and revenue	Are connection scope, commissioning duties,	Delayed energization, curtailment, or payment mismatch	Interface matrix, grid milestone plan, and downside case

	metering, dispatch, and settlement aligned?		
Engineering and construction	Do design, delivered equipment, work quality, schedule, and warranties form one executable system?	Cost overrun, delay, defect, or lost warranty	Site inspection, completion estimate, claims map, and integrated master schedule
Environment and stakeholders	Have impact, biodiversity, land-use, labour, safety, and grievance obligations been addressed?	Delay, liability, lender non-compliance, or loss of social acceptance	Gap assessment, corrective action plan, engagement owner, and monitoring
Finance, tax, and insurance	Do funding, security, tax, insurance, and cash waterfall remain workable after control changes?	Funding gap, uninsured loss, leakage, or default	Sources-and-uses reconciliation, consent map, coverage review, and liquidity reserve
Organization and integration	Can the buyer assume control without interrupting delivery or losing critical knowledge?	Decision paralysis, staff loss, claim prejudice, or reporting failure	Day-one controls, retention plan, delegation schedule, and 100-day integration

4.3 Risk ranking and decision rights

Risk ranking needs defined probability and impact. Impact covers schedule, capital, generation, enforceability, safety, environment, and financeability. Probability reflects event likelihood and evidence quality. A weakly documented low-probability issue is not equivalent to one supported by tested evidence.

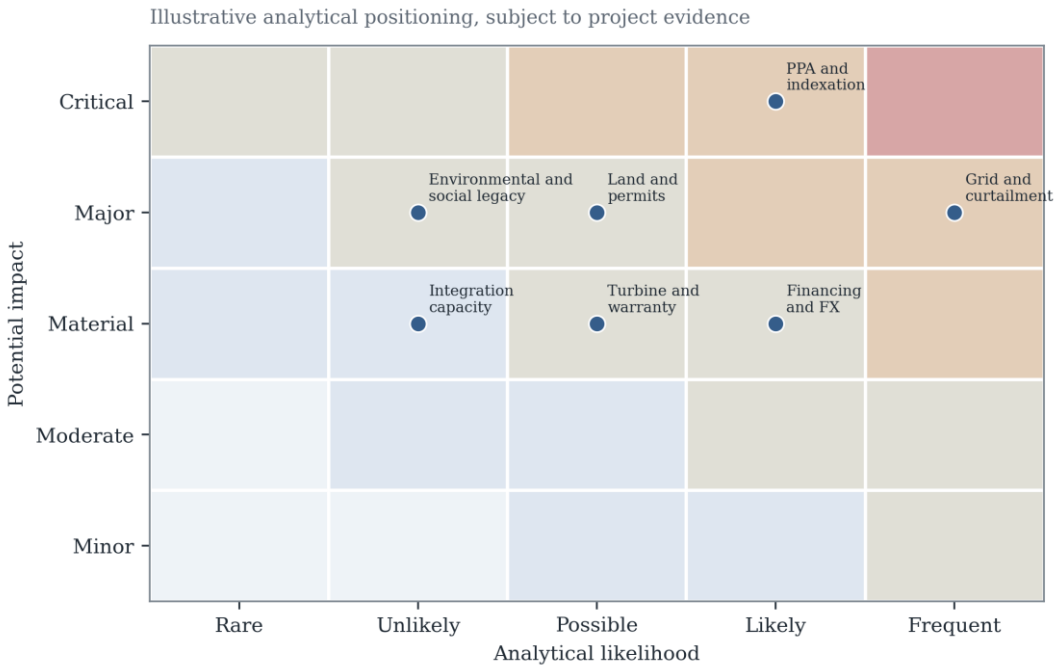


Figure 2. Analytical risk matrix for wind asset acquisition

Source: Developed by the author as a project specific assessment template.

The matrix names the owner before and after closing and the escalation gate. Issues may require correction, a condition, reserve, remedy, or post-closing action. Every accepted risk needs an operational owner.

5. Comparative case evidence

5.1 Zhanatas and Shokpar: build capability with partner continuity

Public facts. EBRD identifies Zhanatas as a 100 MW project owned by China Power International Holding and Visor, with multi-institution financing [18]. AIIB also documents the project [19]. The record supports classification as jointly owned greenfield development.

AIIB describes Shokpar as an adjacent 100 MW project co-developed by CPIH and Visor [20]. EBRD treats it as a separate project [21]. Adjacency supports learning and shared infrastructure, but not legal or environmental equivalence.

Public allegation. EBRD's accountability registry records a Shokpar complaint alleging threatened-bird collision risk and questioning environmental categorization [22]. This article makes no merits finding. The complaint shows why repetition still requires fresh site evidence and a grievance process.

Practitioner interpretation. Organizational learning transfers, but project consent remains local. Procurement knowledge, lender experience, and partner routines can repeat. Biodiversity conclusions, stakeholder maps, and land assumptions cannot. Shokpar required renewed assessment despite stronger complementarity and option value.

5.2 Dostyk: buying time while governing inherited interfaces

Public facts. Reporting attributed to China Export and Credit Insurance Corporation says CPIH acquired Dostyk, 50 MW, during construction. Medium and long-term export buyer's credit insurance reportedly covered 95 percent of both political and commercial risks. It supported RMB project finance from China Construction Bank Astana Branch, with about USD 45 million insured. The source called it Kazakhstan's first renewable project-finance case supported by this insurance form [23]. Public portfolio material also identifies Dostyk at 50 MW [24].

No inference is made here about the acquisition price, seller, transaction timing, loan amount, tenor, pricing, security package, or private allocation of liabilities. Those matters are outside the evidence set.

The structure can reduce lender exposure to covered political and commercial risks and widen RMB financing channels. It does not remove construction, resource, grid, environmental, or operating risk. Insurance neither repairs permits nor guarantees yield.

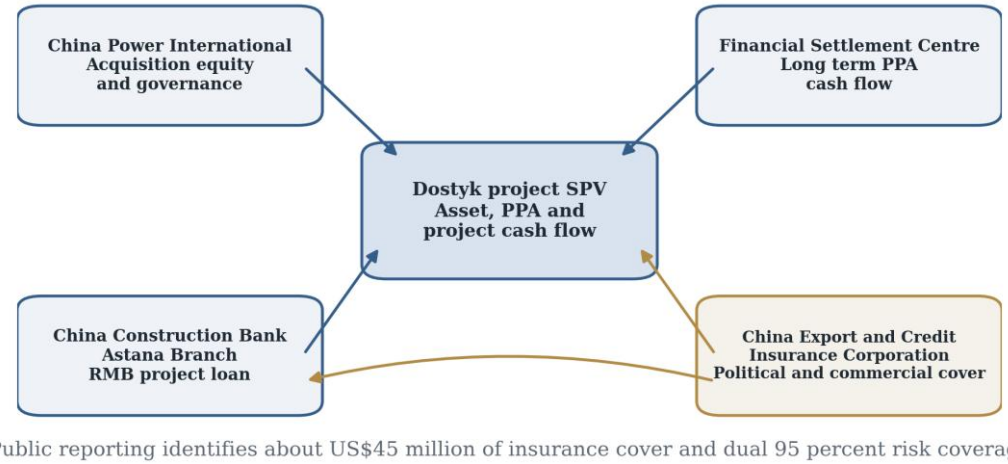


Figure 3. Publicly reported financing and risk allocation structure for the Dostyk project

Source: Compiled from public reporting by China Export and Credit Insurance Corporation.

Practitioner interpretation. Acquisition can save development time but place the buyer at the narrowest project stage. Design may be frozen, equipment delivered, claims forming, and the grid date dominant. The buyer has less freedom than a developer and less evidence than an operating-asset acquirer.

Diligence must therefore test interfaces: design consistency, warranty survival, contractor notices, testing duties, and the completion budget. Finance must fit that completion case. Dostyk shows how insurance and RMB finance can help when evidence and residual risk remain governable.

5.3 Arkalyk: identity before evaluation

The name Arkalyk appears in public sources in connection with two different wind assets. They should not be combined.

Public fact, Arkalyk-Kostanay. A regional-government record identifies the completed plant near Arkalyk city as a 48 MW asset with 10 turbines, names KazWindEnergy as investor, and reports commissioning in December 2023 [25]. KOREM's earlier 2022 auction record describes a 50 MW tendered project in Kostanay Region and names Aspan Energo LLP as the winner [26]. The difference between tendered and completed capacity, together with the named entities, means that any ownership or development sequence must be established through corporate, registry, contract, and coordinate evidence rather than inferred from the project name.

Public fact, Arkalyk-Akmola. AIIB's non-technical summary for the CPIH 220 MW Akmola portfolio describes Arkalyk Wind Power Plant LLP as a separate 15 MW subproject with three 5 MW turbines near Bulaksay and Saryoba, northeast of Astana [27]. It is distinct from the 48 MW plant near Arkalyk city.

Arkalyk-Akmola and Arkalyk-Kostanay are analytical labels, not asserted legal names. The distinction illustrates the first diligence rule: define the object before evaluating it.

Practitioner interpretation. Similar names can attach reports, permits, grid studies, or models to the wrong site. An identity certificate should reconcile entity, registration number, coordinates, region, capacity, grid point, auction record, and ownership before review.

AIIB's separate disclosure of five CPIH plants totaling 220 MW reinforces this need [27]. Portfolio capacity cannot replace an asset-level rights chain.

5.4 The announced 1 GW program: partnership as a governance platform

Public facts. In 2023, Samruk-Kazyna announced cooperation on a 1 GW wind and storage program and stated that the first 500 MW was planned in Zhambyl Region [28]. Official materials in 2026 described a 1 GW wind project in Pavlodar Region [29], while Samruk-Energy reported a construction launch near Ekibastuz with 300 MW of storage and expected investment of approximately USD 1.2 billion [30]. The reviewed documents do not establish legal continuity between every element of the 2023 and 2026 announcements.

Program records must therefore be read chronologically because location, scope, storage, and implementation can evolve. They show partnership direction and increasing development maturity, while each right and milestone still needs to be attached to a specific legal entity, site, and evidence date.

Practitioner interpretation. At 1 GW, partnership is a platform for sequencing land, grid, storage, procurement, industry, finance, and public coordination. Phasing preserves option value through gates, data rights, expansion options, funding calls, and exit rules.

Governance should separate program decisions from project-company control. Reserved matters can cover sites, standards, localization, portfolio finance, and expansion. Project boards control permits, budgets, contracts, construction, and operations. Information, deadlock, and funding-default rules must work within the schedule.

RIGOR indicates high complementarity and option value. Rights vary by site, so common data standards and project-level risk acceptance remain necessary.

6. Findings and propositions

6.1 Entry mode relocates uncertainty

The comparison supports six propositions.

Proposition 1. Greenfield development is preferred when project rights are incomplete but the investor has a superior capability to create, sequence, and document them. The investor accepts time risk to preserve design control.

Proposition 2. Construction-stage acquisition is preferred when rights are substantially formed, evidence is independently auditable, inherited interfaces are governable, and the value of saved time exceeds expected completion and integration costs.

Proposition 3. Joint development is preferred when local or institutional complementarity is material and continuing, and when staged governance can preserve option value without making accountability ambiguous.

Proposition 4. Construction-stage acquisition risk is non-linear. It can peak when physical commitment is high but commissioning evidence and operating cash flow are still absent. Percentage completion alone is therefore an inadequate maturity measure.

Proposition 5. Financing is a mediator between entry mode and performance. Insurance, lender participation, currency, and security can improve fundability, but only after project risks are identified and allocated. Finance does not substitute for evidence.

Proposition 6. Repeat-project learning improves organizational process more reliably than it reduces site-specific environmental and stakeholder risk. Each adjacent or replicated project requires a new local evidence baseline.

These propositions refine the usual build-or-buy choice. Partnership is not a compromise placed between two pure modes. It is a distinct governance technology. Acquisition is not simply a faster greenfield project. It is a transfer of an evolving network of rights and interfaces.

Greenfield development is not absence of legacy. Early choices become the legacy inherited by the operating team.

6.2 A gate-based decision process

The practical process has five gates. Gate 1 confirms project identity and strategic fit. Gate 2 applies RIGOR and compares feasible entry modes. Gate 3 verifies the evidence base and identifies fatal flaws. Gate 4 integrates risk into valuation, financing, contracts, and conditions. Gate 5 approves post-entry governance and assigns every accepted risk to an owner.

At each gate, the investor should retain the option to change mode. A proposed acquisition with unreliable evidence may become a staged partnership. A joint development with weak governance may become a contracted greenfield service arrangement. A greenfield pipeline with mature transferable rights may support a later sell-down or portfolio financing. Mode flexibility is valuable when it is designed in advance. Unplanned mode change usually signals that earlier uncertainty was misclassified.

7. Managerial and policy implications

For investors, the main implication is to select an entry mode and a governance system together. The investment paper should state why the proposed mode is superior to the two alternatives. It should also show which RIGOR dimensions drive that conclusion. A project that is financially attractive under one mode may be ungovernable under another.

For investment committees, evidence quality should be visible. Assumptions should be tagged as verified, independently assessed, management case, or unresolved. Public announcements should not be treated as project rights. Complaints should be recorded as allegations until assessed, but they should not be ignored merely because they are unproven. Practitioner judgment should explain a decision, not silently replace evidence.

For lenders and insurers, the integrated model provides a common bridge to sponsor diligence. A financing structure can allocate defined risks, but the asset must still reach completion and generate cash. Alignment is improved when the lender's conditions, insurer's coverage requirements, construction completion tests, and shareholder integration plan use the same evidence map.

For public partners, program-level cooperation benefits from transparent gates. Large announced capacity should be divided into developable units with clear land, grid, environmental, storage, procurement, and financing milestones. This allows public objectives and investor discipline to reinforce each other. It also makes changes in location or scope traceable rather than contradictory.

For regulators and market institutions, consistent project identifiers and accessible auction, grid, permitting, and ownership information can reduce diligence cost. Lower information friction attracts capital without weakening standards. Standardization is particularly valuable when several assets share similar names or when ownership changes during development.

International standards help define the environmental and social boundary. IFC Performance Standards address assessment, labour, communities, land, biodiversity, and cultural heritage [31]. EBRD policy adds lender-oriented appraisal, engagement, and monitoring [32]. OECD guidance treats due diligence as continuous identification, mitigation, tracking, communication, and remediation [33]. These standards do not replace Kazakhstan law, but they add a financeability lens.

8. Limitations and future research

This study has four limitations. First, it relies on public evidence. Public records are uneven across projects and may lag current status. Second, the case set is analytically selected and cannot establish statistical causality. Third, confidential transaction economics, internal negotiations, seller information, acquisition dates, and debt terms are deliberately excluded. The paper therefore evaluates governance logic rather than transaction returns. Fourth, practitioner interpretation can add process insight but may also introduce perspective bias. The explicit evidence labels reduce this risk but cannot eliminate it.

Future research could test the RIGOR dimensions across a larger sample of Kazakhstan renewable projects. Interviews with developers, lenders, insurers, regulators, and local stakeholders could examine whether the framework predicts mode changes and schedule outcomes. Quantitative work could model the option value of staged entry under auction, grid, and currency uncertainty. A longitudinal study could also compare how due diligence findings are converted into post-closing action and operating performance.

9. Conclusion

The choice to build, buy, or partner is a choice about the timing, ownership, and governability of uncertainty. Kazakhstan's wind market illustrates all three routes. Zhanatas and Shokpar show the value of repeated development capability and partner continuity, while also showing why site-specific diligence must be renewed. Dostyk shows the speed and finance potential of a construction-stage acquisition, together with the need to govern inherited completion interfaces. The two public Arkalyk records show that project identity is a substantive control. The announced 1 GW program shows why partnership at scale requires a phased governance platform.

The RIGOR framework connects rights completeness, information reliability, governance complementarity, option value, and residual risk to entry choice. The integrated acquisition model then links specialist diligence to loss pathways, liability allocation, and post-closing leadership. Together, they move the decision beyond a simple comparison of price and speed.

The practical conclusion is concise. Build when development capability creates control. Buy when evidence makes inherited time governable. Partner when complementarity and option value exceed the costs of shared control. In every mode, treat evidence as an operating asset. Capital can enter through a transaction. Sustainable value enters through governance.

References:

1. Williamson, O. E. (1985). The Economic Institutions of Capitalism. New York: Free Press.

2. Anderson, E., and Gatignon, H. (1986). Modes of foreign entry: A transaction cost analysis and propositions. Journal of International Business Studies, 17(3), 1-26.

3. Hennart, J. F., and Park, Y. R. (1993). Greenfield versus acquisition: The strategy of Japanese investors in the United States. Management Science, 39(9), 1054-1070.

4. Brouthers, K. D., and Brouthers, L. E. (2000). Acquisition or greenfield start-up? Institutional, cultural and transaction cost influences. Strategic Management Journal, 21(1), 89-97.

5. Kogut, B. (1991). Joint ventures and the option to expand and acquire. Management Science, 37(1), 19-33.

6. Dixit, A. K., and Pindyck, R. S. (1994). Investment Under Uncertainty. Princeton: Princeton University Press.

7. Trigeorgis, L., and Reuer, J. J. (2017). Real options theory in strategic management. Strategic Management Journal, 38(1), 42-63.

8. Meyer, K. E., and Estrin, S. (2001). Brownfield entry in emerging markets. Journal of International Business Studies, 32(3), 575-584.

9. Slangen, A. H. L., and Hennart, J. F. (2007). Greenfield or acquisition entry: A review of the empirical foreign establishment mode literature. Journal of International Management, 13(4), 403-429.

10. Haleblian, J., Devers, C. E., McNamara, G., Carpenter, M. A., and Davison, R. B. (2009). Taking stock of what we know about mergers and acquisitions: A review and research agenda. Journal of Management, 35(3), 469-502.

11. Zollo, M., and Meier, D. (2008). What is M&A performance? Academy of Management Perspectives, 22(3), 55-77.

12. Wustenhagen, R., and Menichetti, E. (2012). Strategic choices for renewable energy investment: Conceptual framework and opportunities for further research. Energy Policy, 40, 1-10.

13. Gatzert, N., and Kosub, T. (2016). Risks and risk management of renewable energy projects: The case of onshore and offshore wind parks. Renewable and Sustainable Energy Reviews, 60, 982-998.

14. Steffen, B. (2018). The importance of project finance for renewable energy projects. Energy Economics, 69, 280-294.

15. Republic of Kazakhstan. (2009, as amended). Law No. 165-IV on Support for the Use of Renewable Energy Sources. Adilet Legal Information System. https://adilet.zan.kz/eng/docs/Z090000165_

16. Financial Settlement Centre for Support of Renewable Energy Sources. Renewable energy auction materials and sector information. https://rfc.kz/en/res-sector/auctions/materials/

17. International Energy Agency. (2022). Kazakhstan 2022: Energy Sector Review. Paris: IEA. https://www.iea.org/reports/kazakhstan-2022

18. European Bank for Reconstruction and Development. Zhanatas Wind Power Plant, project 50569. https://www.ebrd.com/home/work-with-us/projects/psd/50569.html

19. Asian Infrastructure Investment Bank. (2019). Kazakhstan: Zhanatas 100 MW Wind Power Plant. https://www.aiib.org/en/projects/details/2019/approved/Kazakhstan-Zhanatas-100-MW-Wind-Power-Plant.html

20. Asian Infrastructure Investment Bank. (2023). Kazakhstan: Shokpar 100 MW Wind Power Project. https://www.aiib.org/en/projects/details/2023/approved/Kazakhstan-Shokpar-100MW-Wind-Power-Project.html

21. European Bank for Reconstruction and Development. Shokpar Wind Power Project, project 52946. https://www.ebrd.com/home/work-with-us/projects/psd/52946.html

22. European Bank for Reconstruction and Development, Independent Project Accountability Mechanism. KAZREF II Shokpar Wind, case registry. https://www.ebrd.com/home/what-we-do/projects/independent-project-accountability-mechanism/case-registry/KAZREF-II-Shokpar-Wind.html

23. China Export and Credit Insurance Corporation, reproduced by Asian Financial Cooperation Association. China Sinosure underwrites Kazakhstan Dostyk 50 MW wind power

project and supports project financing for a Chinese enterprise acquisition. https://cn.afca-asia.org/Portal.do?contentID=3721&method=detailView&returnChannelID=3

24. SANY Renewable Energy. Global project portfolio and Kazakhstan Dostyk 50 MW project. https://en.sanyre.com.cn/upload/1/cc/1755655571515.pdf

25. Akimat of Kostanay Region. (2024, February 1). Energy supply and commissioning of the 48 MW Arkalyk wind plant. https://www.gov.kz/memleket/entities/kostanai/press/news/details/697336?lang=ru

26. Kazakhstan Electricity and Power Market Operator. (2022). Results of the auction for selection of renewable energy projects, Arkalyk wind project, 28 November 2022. https://www.korem.kz/eng/press-centr/archive_otrasli?cid=126&rid=11046

27. Asian Infrastructure Investment Bank. (2024). Akmola Wind Power Project Non-Technical Summary, including the 15 MW Arkalyk Wind Power Plant LLP subproject. https://www.aiib.org/en/projects/details/2024/_download/Kazakhstan/1-Akmola-Project-Non-Technical-Summary.pdf

28. Samruk-Kazyna. (2023, December 6). Samruk-Kazyna and China Power International Holding advance a 1 GW wind project with energy storage. https://www.sk.kz/press-center/news/77426/?lang=en

29. Samruk-Kazyna. (2026, January 29). Kazakhstan and China advance a 1 GW wind project in Pavlodar Region. https://sk.kz/press-center/news/78809/?lang=en

30. Samruk-Energy. (2026, March 30). Samruk-Energy launches construction of a major wind power project in the CIS. https://samruk-energy.kz/en/press-center/news/2664-samruk-energy-jsc-launches-construction-of-another-major-wind-power-project-in-the-cis

31. International Finance Corporation. (2012). Performance Standards on Environmental and Social Sustainability. Washington, DC: IFC. https://www.ifc.org/en/insights-reports/2012/ifc-performance-standards

32. European Bank for Reconstruction and Development. (2019). Environmental and Social Policy. London: EBRD. https://www.ebrd.com/news/publications/policies/environmental-and-social-policy-esp.html

33. Organisation for Economic Co-operation and Development. (2018). OECD Due Diligence Guidance for Responsible Business Conduct. Paris: OECD Publishing. https://doi.org/10.1787/15f5f4b3-en

UDC 620.91:621.311(574)

Wang Liang
DBA Student
Al-Farabi Kazakh National University
(Almaty, Kazakhstan)

FROM CAPACITY EXPANSION TO SYSTEM VALUE: COORDINATED DEVELOPMENT OF RENEWABLE ENERGY, GRID FLEXIBILITY AND ENERGY STORAGE IN KAZAKHSTAN

Abstract. By the end of 2024, Kazakhstan had connected 3,038.6 MW of renewable capacity, which produced 7.5551 TWh or 6.4% of national generation. Consumption nevertheless exceeded domestic generation by 2.0801 TWh, and installed generation exceeded available capacity by 6,423.9 MW. These indicators show why the next stage should evaluate projects by power-system contribution as well as nameplate capacity and auction tariff. Using official secondary data, policy documents and a structured system-value framework, this paper examines adequacy, flexibility, transmission, balancing, storage and localization. The 2026 auction schedule represents progress because it differentiates technologies and zones and introduces battery requirements for selected solar lots. Coordinated planning remains necessary because renewable value can fall when output is concentrated in congested locations or hours with insufficient reserves. Four measures are proposed: locational procurement, anticipatory transmission, technology-neutral flexibility markets and phased localization tied to lifecycle capability. The analysis also distinguishes MW from MWh, the older 10% benchmark from the later 15% renewable-share instruction for 2030, and ordinary 20-year auction PPAs from project-specific 25-year arrangements. The framework supports a reliable, investable and locally beneficial transformation.

Keywords: Kazakhstan; renewable energy; system value; grid flexibility; energy storage; auctions; localization

JEL codes: Q42; Q48; L94; O33

1. Introduction

Kazakhstan has strong wind and solar resources, a renewable-energy support law, competitive auctions and long-term decarbonization objectives. At the end of 2024, 157 renewable facilities with 3,038.6 MW of capacity supplied 6.4% of national generation [6]. The 2026 auction schedule offers another 915 MW, comprising 450 MW of wind, 265 MW of solar and 200 MW of hydropower [9]. Capacity expansion remains necessary, although commissioned MW alone no longer provides a sufficient measure of success.

The operational context is demanding. Consumption reached 119.9955 TWh in 2024, compared with domestic production of 117.9154 TWh, and the 2.0801 TWh deficit was covered through net exchanges with Russia and Central Asia [6]. At the beginning of 2025, installed generation was 25,314.2 MW but available capacity was 18,890.3 MW [7]. Maintenance, fuel, technical and resource constraints explain this difference and demonstrate that installed MW, available MW and annual MWh represent different services.

This issue is especially important for variable renewable energy. A wind or solar plant contributes low-carbon electricity, fuel-price protection and investment, but its production depends on weather and time. When many projects produce simultaneously, their marginal market value can decline; when they are far from demand, network cost rises; and when forecast

errors increase, the system requires additional balancing resources [18,19]. The relevant policy question thus concerns the composition, location and operating capability of new capacity, together with the infrastructure and market arrangements that allow it to serve consumers reliably.

The objective of this paper is to develop a system-value approach for the coordinated development of renewable energy, grid flexibility and electricity storage in Kazakhstan. Three research questions are addressed. First, what do the official 2024 data reveal about the relationship among installed capacity, available capacity, energy production and cross-border balancing? Second, how does the 2026 auction schedule begin to incorporate location and storage, and what gaps remain? Third, which policy and investment measures can increase reliability, affordability and domestic industrial benefits during the next phase of deployment?

The paper applies system-value concepts to current official Kazakh data, integrates generation, networks, balancing and industrial policy, and converts the diagnosis into sequenced measures. Because annual secondary data cannot reproduce hourly dispatch, probabilistic adequacy or optimal storage sizing, the framework also identifies information required for subsequent project appraisal.

2. Literature and Policy Review

2.1 From plant cost to power-system value

Levelized cost compares lifetime average technology costs but does not fully represent a plant's contribution to an interconnected system. The market value of variable renewables depends on the temporal and spatial correlation between output and prices, and coincident production can reduce the value factor as penetration grows [18]. System LCOE adds profile, balancing and grid-related costs to plant cost [19]. A low bid tariff therefore supports procurement efficiency, while consumer value also depends on when, where and how power is delivered.

System value has six connected dimensions: **energy value**, measured by useful MWh and displaced fuel; **capacity value**, meaning dependable contribution during scarcity; **flexibility value**, including ramps, reserves and frequency response; **network value**, determined by location, losses and congestion; **environmental and risk value**, including emissions and fuel-price exposure; and **industrial value**, covering domestic services, skills and maintenance. Joint assessment is necessary because performance differs across dimensions.

Flexibility manages supply-demand variability across seconds, hours, days and seasons at reasonable cost [15]. It can come from flexible generation, interconnection, demand response, storage, forecasting, networks and balancing. Batteries respond quickly, but their role depends on power, duration, cycling, location and dispatch rights [16,20]. Wind, solar and hydropower may also be complementary, subject to testing with synchronized resource and demand data [21].

International experience shows that auctions can incorporate connection readiness, location, delivery schedules and grid support if requirements are transparent and competition is preserved [17]. Unvalued obligations raise bids or reduce participation, while socialized grid costs without locational signals can produce sites that are cheap for the plant and costly for the system.

2.2 Kazakhstan's policy trajectory and target scopes

Kazakhstan's renewable-energy framework has evolved through several layers. The Law on Support of the Use of Renewable Energy Sources defines the principal support mechanism, while the Electric Power Industry Law establishes the wider sectoral and market setting [4,5].

The Green Economy Concept created a phased long-term direction and is widely restated as a renewable share in electricity production of 3% in 2020, 6% in 2025 and 10% in 2030, followed by 50% from alternative and renewable sources in 2050 [1]. The 2050 indicator has a broader technology scope than the renewable-only milestones and should be interpreted accordingly.

In May 2021, the President instructed the government to raise the share of renewable energy in electricity generation to 15% by 2030, explicitly identifying this as an increase from the earlier 10% goal [2]. The two figures should not be added, averaged or treated as simultaneous definitions of the same target. The 10% figure is the older benchmark embedded in the Green Economy policy trajectory; the 15% figure is the later presidential policy instruction for the same year and a more ambitious renewable-generation share. Separate from both, Kazakhstan's 2060 Carbon Neutrality Strategy establishes economy-wide decarbonization and states that long-term renewable development will be accompanied by electricity-storage systems [3]. Likewise, the strategy's 15% unconditional greenhouse-gas reduction by 2030 is measured against 1990 emissions and is a different indicator from a 15% renewable share in electricity generation.

Auctions introduced in 2018 improved price discovery and investor access, while long-term offtake and indexation supported bankability. The 2023 single-buyer and balancing reforms increased the importance of schedules and imbalance settlement [7]. Research emphasizes regulatory stability, spatial constraints, legacy thermal generation, combined heat-and-power operations and transmission limits [22-25]. Scenario work confirms that grids, flexibility and technology coordination affect the cost of deeper integration [26].

Policy has begun to move toward integrated projects: designated 2026 solar lots require batteries, while grid projects integrate the west and reinforce the south [8]. The interfaces among auctions, transmission, balancing and localization nevertheless require further definition because a plant auction alone cannot determine the least-cost system portfolio.

3. Data and Method

3.1 Data sources and unit discipline

The study uses secondary sources available as of August 2026. The principal quantitative source is KEGOC's 2024 Annual Report, including the national electricity balance, generation structure and investment program [6-8]. Auction information is taken from the official 2026 announcement, the published auction calendar and official KOREM conditions [9-11]. Policy interpretation relies on laws, presidential documents and the Carbon Neutrality Strategy [1-5]. International comparison and conceptual definitions are drawn from the IEA, IRENA and peer-reviewed literature [14-21].

MW measures instantaneous power, whereas **MWh** measures energy delivered or stored. A 30 MW battery with 120 MWh of usable energy has four hours of nominal full-power duration. A disclosed 600 MWh energy capacity reveals neither the MW rating nor duration. Annual electricity is reported in TWh or GWh, and plant or network capability in MW or MVA.

3.2 Structured system-value assessment

The analytical framework evaluates a renewable project or portfolio (j) through six value channels and five enabling conditions. In conceptual form:

$$SV_j = EV_j + CV_j + FV_j + NV_j + ERV_j + IV_j - IC_j,$$

where (EV) is useful energy value, (CV) is dependable capacity value, (FV) is flexibility value, (NV) is network and locational value, (ERV) is environmental and risk value, (IV) is industrial value, and (IC) represents integration costs already not captured in the other terms. This expression is an accounting framework, not an estimated monetary equation in the present

paper. Monetization would require hourly prices, scarcity probabilities, nodal constraints, reserve procurement, curtailment, outage distributions and technology cost data that are not fully public.

The assessment proceeds in four steps: (1) diagnose the 2024 energy and capacity baseline; (2) map renewable additions against zones, connection arrangements and flexibility requirements; (3) test whether auction and market rules reward the relevant services; and (4) identify a staged set of institutional, infrastructure and industrial measures. The framework avoids double counting by treating a service as value only once and by separating project benefits from costs transferred to the grid or consumers.

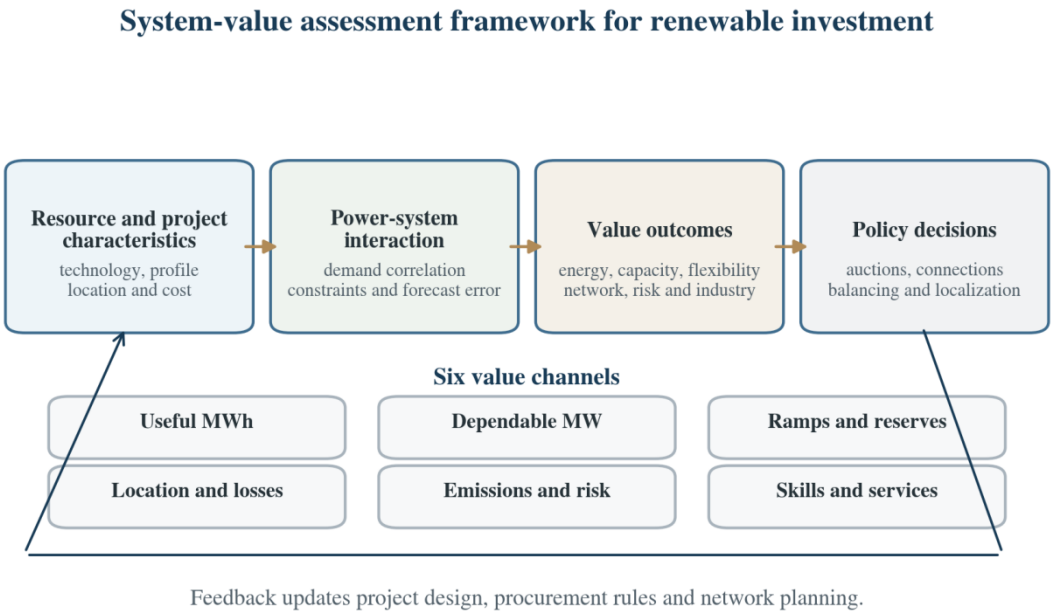


Figure 1. System-value assessment framework for renewable investment

Source note: Author's design based on the system-value literature [15,18,19] and auction-design guidance [17].

4. Kazakhstan's 2024 Power-System Baseline

4.1 Capacity, availability and energy

Table 1 presents the indicators most relevant to system value. Kazakhstan's 25,314.2 MW of installed capacity exceeded available capacity by 6,423.9 MW, meaning that 74.6% of nameplate capacity was categorized as available at the beginning of 2025. This ratio should not be interpreted as a universal plant-performance measure because availability definitions differ by technology. KEGOC's reported methodology credits 30% of run-of-river hydropower capacity, 20% of wind and 0% of solar in the available-capacity balance [7]. The methodology is conservative for adequacy planning and makes clear why nameplate additions cannot be directly substituted for dependable capacity.

Table 1. Selected indicators of Kazakhstan's electricity system, 2024

Indicator	Official value	Derived interpretation
Total installed generation at 1 January 2025	25,314.2 MW	Nameplate capability across 233 plants
Available generation	18,890.3 MW	74.6% of installed capacity; gap of 6,423.9 MW
Renewable installed capacity	3,038.6 MW	12.0% of total installed capacity
Wind / solar / small hydro / bioenergy capacity	1,525.7 / 1,216.6 / 295.2 / 1.1 MW	Technology mix within reported RES fleet
National generation	117.9154 TWh	Increase of 4.5% year on year
National consumption	119.9955 TWh	Increase of 4.3% year on year
Generation-consumption balance	-2.0801 TWh	Domestic annual energy deficit
Renewable generation	7.5551 TWh	6.4% of national generation; increase of 12.5%
Wind / solar / small hydro / bioenergy generation	4.4976 / 1.8956 / 1.1613 / 0.0006 TWh	Implied annual capacity factors of about 33.7%, 17.8%, 44.9% and 6.2%
Net exchange with Russia	+3.4111 TWh	Imports exceeded exports
Net exchange with Central Asia	-1.3310 TWh	Exports exceeded imports

Source note: KEGOC 2024 Annual Report [6,7]. Percentages, capacity gaps, net exchanges and implied capacity factors are the author's calculations; a positive exchange indicates net imports into Kazakhstan.

Wind supplied 59.5% of reported renewable generation, solar 25.1% and small hydro 15.4%. The derived annual capacity factors shown in Table 1 indicate resource performance, but peak contribution requires several years of coincident hourly output, demand and outage data.

4.2 Geography, interconnection and balancing

Location is central because northern and central generation, southern load and solar potential, and the historically separate western zone create different network conditions. An annual national balance can coexist with regional scarcity, congestion or difficult ramps.

Cross-border exchanges provided a material balancing and energy function in 2024. Imports from Russia were 4.9260 TWh and exports to Russia were 1.5149 TWh, producing net imports of 3.4111 TWh. Kazakhstan simultaneously recorded net exports of 1.3310 TWh to Central Asia. The combined net position equals the 2.0801 TWh domestic generation deficit [6]. Interconnection creates economic and operational value, but long-term adequacy should not depend on unpriced or uncertain external flexibility. The IEA has previously observed that

Kazakhstan has limited flexible generation and uses parallel operation with Russia to manage imbalances and frequency, a challenge that increases as variable output grows [14].

KEGOC's investment program directly addresses several spatial constraints. The Western Kazakhstan integration project includes the 500 kV Karabatan-Ulke line of about 604.3 km and is intended to connect the western zone to the national system, thereby allowing the flexible capacity of the west to support the wider system. Southern-network reinforcement includes new lines and substation upgrades, while planned reconstruction covers thousands of kilometers of 220 kV and 500 kV assets [8]. These investments should be evaluated with renewable-auction decisions because a renewable plant's commissioning date and its connection asset form one delivery chain.

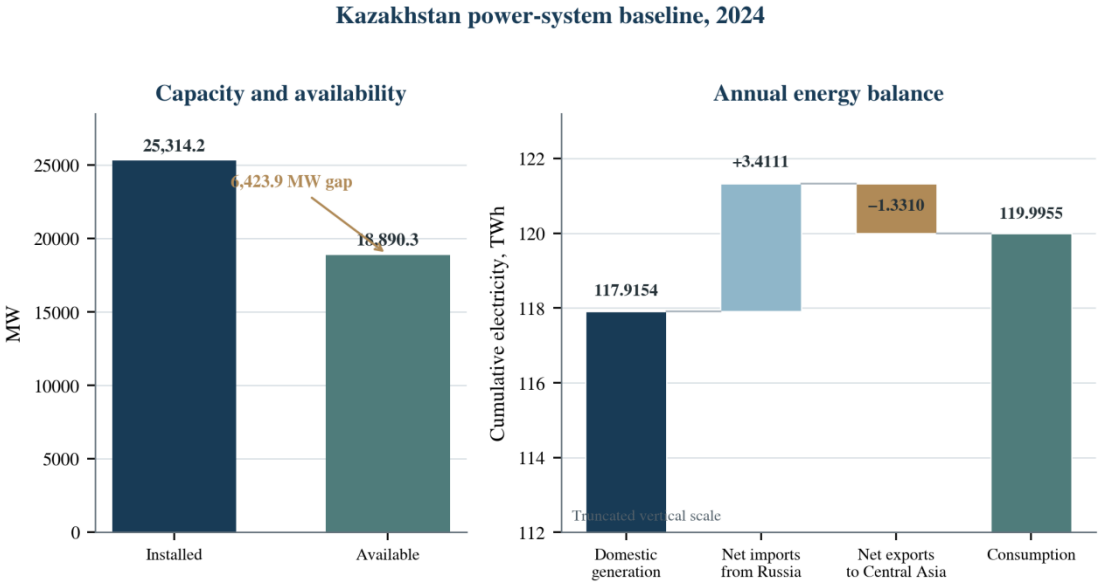


Figure 2. Kazakhstan power-system capacity and annual energy baseline, 2024

Source note: Author's visualization specification using official KEGOC data [6,7].

5. The 2026 Auction Schedule and the Emerging Role of Storage

5.1 What the schedule contributes

The Ministry of Energy approved a 915 MW renewable-auction program for 2026. The program is differentiated by technology and zone and combines open auctions, in which bidders may bring their own land and connection point, with auctions linked to specified zones or connection arrangements [9,10]. This differentiation is meaningful. It can reduce speculative project selection, align capacity with regional needs and reveal the value of prepared grid access.

Table 2. Kazakhstan's 2026 renewable-energy auction schedule

Date in 2026	Technology	Zone or region	Offered capacity	Published commissioning period	Storage condition	Ordinary PPA term
27 April	Wind	Western zone, Atyrau	250 MW	36 months	No explicit battery condition reported	20 years
28 April	Solar	Southern zone, Turkestan	50 MW	24 months	Battery requirement applies under solar open-auction conditions	20 years

29 April	Solar	Almaty region	100 MW	24 months	Battery power at least 30% of plant MW and energy for at least 4 hours	20 years
30 April	Solar	Almaty region	50 MW	24 months	Same minimum battery condition	20 years
18 May	Solar	Kyzylorda region	65 MW	24 months	Same minimum battery condition	20 years
19 May	Wind	Northern zone	100 MW	36 months	No explicit battery condition reported	20 years
20 May	Wind	Northern zone	100 MW	36 months	No explicit battery condition reported	20 years
21 May	Hydropower	All zones	150 MW	60 months	No explicit battery condition reported	20 years
22 May	Hydropower	All zones	50 MW	60 months	No explicit battery condition reported	20 years
Total			915 MW			

Source note: Official aggregate program and support terms [9]; detailed calendar and regions [10]; official KOREM solar-auction storage condition [11]. "No explicit battery condition reported" means that the cited schedule did not state one and does not prove that no other technical obligations apply.

5.2 Interpreting the battery requirement

The storage condition in selected solar auctions illustrates why MW and MWh must remain distinct. The requirement states that battery power should be at least 30% of the solar plant's installed MW and that stored energy should sustain this power for at least four hours [11]. For a 100 MW solar plant, the minimum implied battery specification is 30 MW and 120 MWh. For 50 MW of solar, it is 15 MW and 60 MWh; for 65 MW, it is 19.5 MW and 78 MWh. These are minimum arithmetic implications. Usable energy, degradation reserve, temperature derating, auxiliary consumption and grid-code performance may require a larger nominal design.

Such a battery can shift midday output, reduce ramps, improve schedules and provide fast response. The highest-value service still depends on location. Contracts should define dispatch control, state-of-charge management, grid charging, loss settlement, degradation compensation and whether balancing revenue can be combined with the PPA.

Large project-specific arrangements require equally careful unit and contract interpretation. TotalEnergies' Mirny project combines a 1 GW wind plant with 600 MWh of battery energy and a 25-year power-purchase agreement [12]. The disclosed 600 MWh is an energy quantity; absent a disclosed MW rating, battery duration cannot be inferred. The 25-year term is linked to a specific large-scale project arrangement and should not be presented as the ordinary auction rule. The standard 2026 framework provides 20-year PPAs [9]. Maintaining this distinction is important for investor expectations, procurement comparability and public communication.

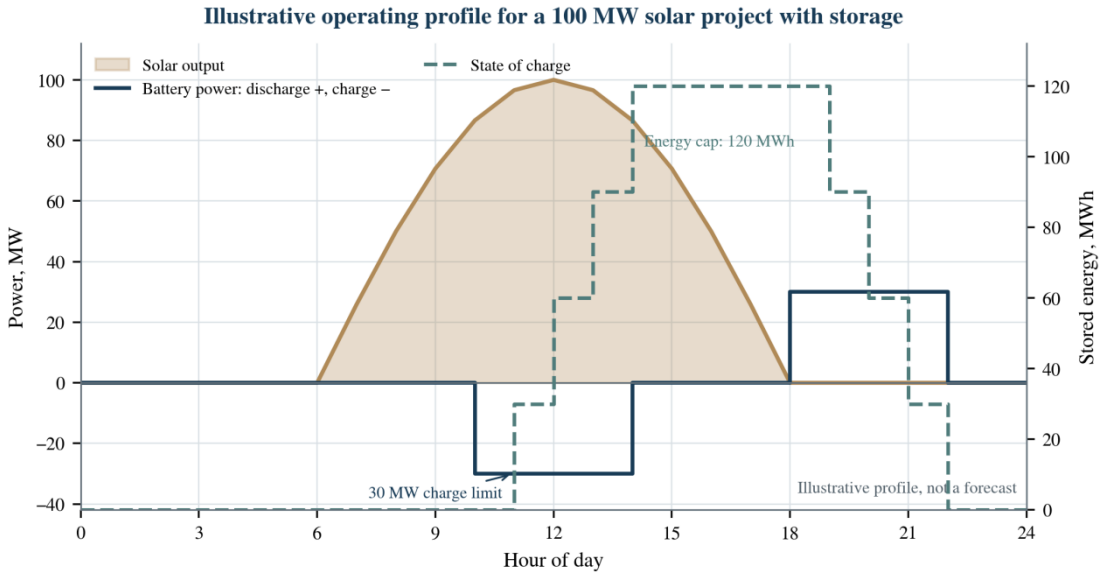


Figure 3. Illustrative operating profile for a 100 MW solar project with storage

Source note: Author's conceptual specification based on the minimum solar-auction storage ratio reported by KOREM [11] and battery-system principles summarized by the IEA [16].

6. A Coordinated Development Strategy

6.1 Procure renewable energy according to value and location

Competitive auctions should add transparent system-value signals. A multi-year procurement map should identify technology, zone, connection point, delivery window and system need, allowing lots to be classified as energy, peak-support, congestion-relief or flexibility products. Price competition would continue within each defined product.

Later, a published value-adjusted tariff could reflect losses, system-funded connection cost, delivery profile and committed services. In the near term, separate locational products are simpler and auditable. Prepared-site auctions can serve strategic corridors, while open auctions can discover developer-led sites [9]. Both require aligned milestones for land, permits, network studies, financing and equipment delivery.

6.2 Treat transmission as an anticipatory transition asset

Transmission should be planned before congestion delays investment. Generation and network projects need integrated schedules and named responsibility for critical interfaces. Priorities are to complete western integration, synchronize southern reinforcement with solar and demand growth, and reconstruct aging 220 kV and 500 kV corridors with justified digital protection and monitoring [8].

Risk allocation should distinguish dedicated connection assets, normally paid by the project, from shared reinforcement recovered through the regulated network. Locational auction signals can maintain discipline. Curtailment rules should separately address grid unavailability, project non-compliance and emergency system action.

6.3 Build a portfolio of flexibility and balancing services

Flexibility procurement should define services, such as frequency response, reserves, ramps and multi-hour energy, and permit competition among batteries, hydropower, flexible thermal units, demand response and interconnection. Renewable producers and aggregators also need timely data, intraday schedule updates and bankable imbalance settlement. Aggregation can reduce individual forecast error.

Storage should be procured by use case: short duration for fast response, four hours for solar shifting, and longer duration for extended adequacy. Every project should disclose rated MW, usable MWh, duration, efficiency, availability, degradation, connection and control rights, together with safety, climate performance, cybersecurity and end-of-life plans. KEGOC's earlier storage feasibility work can be updated with current costs and balancing data [13].

6.4 Link localization to capability, quality and scale

Localization can create employment and maintenance capability, but poorly sequenced mandates can raise tariffs or create shallow assembly. Initial emphasis should be on civil works, structures, cabling, substations, logistics and operations. Subsequent phases can deepen engineering, digital monitoring, equipment repair, battery diagnostics and testing, followed by component production only where demand, quality assurance, export potential and recycling justify scale [2].

Auction documents should request a localization and skills plan rather than rely solely on a percentage. Verifiable commitments may cover domestic expenditure, training, supplier development, local engineering, warranties and lifecycle service centers, announced early enough for credible partnerships with Kazakh firms and universities.

Table 3. Proposed policy matrix for coordinated system-value development

Area	Near-term action, 2026-2028	Medium-term action, 2029-2035	Principal indicator
Renewable procurement	Publish zonal products, connection data and delivery windows; retain price competition	Introduce value-adjusted evaluation informed by hourly and nodal analysis	Useful delivered MWh, curtailment, connection readiness
Transmission	Integrate auction and KEGOC schedules; complete priority western and southern projects	Develop anticipatory corridors and digital network capability	Transfer capacity, congestion hours, project delay
Balancing and flexibility	Define products for reserves, ramps and fast response; improve data access	Enable aggregation, demand response and multi-service storage participation	Reserve sufficiency, imbalance volume and cost
Storage	Specify MW, usable MWh, duration and dispatch rights for every project	Procure multiple durations by system need and location	Availability, cycles, response, avoided curtailment
Localization	Focus on construction, engineering, O&M, training and verifiable suppliers	Deepen repair, testing, components and recycling where scale is proven	Domestic value added, certified personnel, lifecycle capability
Governance and data	Publish auction, grid and balancing performance dashboards	Conduct periodic integrated resource and network planning	Timeliness, data completeness, forecast accuracy

Source note: Author's recommendations based on the diagnostic results in Sections 4 and 5, Kazakhstan's policy objectives [2,3,8,9] and international guidance [15-17].

7. Discussion

The proposed transition from capacity expansion to system value has several implications. First, it changes the meaning of a successful auction. A low tariff remains essential

because consumers should receive the benefit of competition, but success also requires on-time grid connection, predictable output, manageable balancing exposure and durable equipment performance. Procurement reports should therefore track delivery, curtailment, forecast error, local capability and network readiness in addition to the winning tariff and awarded MW.

Second, the framework distributes responsibility more clearly. Developers are responsible for compliant assets, credible forecasts, agreed grid services and delivery milestones. KEGOC is responsible for transparent connection studies, system operation, transmission planning and reliable data. The single buyer and balancing settlement institutions are responsible for bankable contracts and accurate settlement. The Ministry of Energy is responsible for aligning targets, auction volumes, capacity adequacy and industrial policy. Regional authorities, universities and suppliers have roles in land, permitting, skills and local economic development. Coordination requires a formal process because no individual participant can optimize the whole system.

Third, investment certainty must be preserved while rules evolve. Kazakhstan's 20-year ordinary auction PPA is a core bankability instrument [9]. Additional performance requirements should be standardized before bidding and supported by contractual definitions. Large strategic projects may use special arrangements, including a 25-year term such as the Mirny agreement [12], when justified by scale, storage, investment structure and negotiated obligations. Such exceptions should be communicated as project-specific and assessed for value, since investors otherwise may assume that the special term is generally available or that ordinary projects face comparable obligations.

Fourth, storage should be valued through the service it provides. A mandatory ratio is a practical early-stage instrument and can accelerate learning. Over time, however, storage quantity should be derived from adequacy, ramping, congestion and reserve needs. A four-hour battery paired with solar provides a specific daily-shifting capability; it does not solve seasonal deficits, prolonged low-wind periods or all transmission constraints. Conversely, a smaller fast-response battery may create significant frequency value even if its stored MWh are limited. Technology-neutral service markets and location-specific procurement can reveal these differences.

Fifth, localization should be connected to long-term operating quality. Wind turbines, inverters, transformers and batteries require decades of inspection, software support, spare parts and skilled maintenance. Domestic capability in these functions can be more economically valuable than nominal assembly that remains dependent on imported engineering and warranties. For Chinese and other international investors, joint development with Kazakh enterprises and educational institutions can combine global supply-chain scale with local execution, environmental knowledge and workforce development. This model is consistent with Kazakhstan's investment objectives and can improve project resilience.

The analysis also highlights a data agenda. Public hourly demand, zonal generation, curtailment, congestion, reserve activation, imbalance prices and outage statistics would allow researchers and investors to estimate marginal system value. Standardized interconnection queues and hosting-capacity maps would reduce duplicated development expenditure. Storage registers should report MW and MWh separately. These improvements would support the spatial and dispatch modeling already developed for Kazakhstan [25,26] and permit policy to move from qualitative screening to quantitative optimization.

8. Conclusion, Policy Implications and Limitations

Kazakhstan has made substantial progress in renewable-energy development. The 2024 fleet of 3,038.6 MW generated 7.5551 TWh and supplied 6.4% of national electricity. The next phase will occur in a system that also recorded a 2.0801 TWh annual domestic energy deficit and a 6,423.9 MW difference between installed and available generating capacity. These conditions strengthen the case for renewable expansion and require that new projects be coordinated with dependable capacity, grid development and operational flexibility.

The 2026 auction program provides a constructive foundation. Its zonal and technology differentiation can improve portfolio composition, while battery requirements for selected solar lots introduce dispatch capability into procurement. The framework should now be deepened through four measures: procure clearly defined locational products; treat transmission as an anticipatory transition asset; establish transparent markets for balancing and flexibility services; and implement phased localization based on verifiable capability and lifecycle quality.

Three distinctions must remain explicit in policy and investment practice. MW measures power and MWh measures energy, so a storage project requires both values for meaningful evaluation. The older 10% renewable benchmark for 2030 and the later 15% policy instruction refer to alternative ambition levels for a renewable share in electricity generation, while other 15% climate indicators may have entirely different denominators. Ordinary 2026 auction projects receive 20-year PPAs, whereas the cited 25-year term belongs to a special large-project arrangement and is not the general auction standard.

This study is limited by its reliance on annual official statistics and publicly available program documents. It does not estimate hourly curtailment, effective load-carrying capability, probabilistic loss of load, nodal prices, optimal battery duration or the consumer cost of proposed measures. Annual data also cannot capture seasonal hydrology or coincident wind and solar events. Future research should combine at least three to five years of hourly weather, demand, outage and network data in a co-optimized generation-transmission-storage model, test alternative auction and balancing designs, and assess distributional effects on households, industry and regions. Subject to these limitations, the system-value framework offers an implementable direction: renewable capacity, grids, flexibility, storage and local capability should advance as one coordinated investment program.

References:

1. International Energy Agency. “Concept for the Transition of the Republic of Kazakhstan to a Green Economy.” IEA Policies Database, updated 24 January 2022. https://www.iea.org/policies/2138-concept-for-the-transition-of-the-republic-of-kazakhstan-to-a-green-economy

2. President of the Republic of Kazakhstan. “The Head of State Held a Meeting on the Development of the Electric Power Industry.” 26 May 2021. https://akorda.kz/en/the-head-of-state-held-a-meeting-on-the-development-of-the-electric-power-industry-2641630

3. President of the Republic of Kazakhstan. Decree No. 121 of 2 February 2023, “Strategy for Achieving Carbon Neutrality of the Republic of Kazakhstan until 2060.” Adilet Legal Information System. https://adilet.zan.kz/rus/docs/U2300000121

4. Republic of Kazakhstan. Law No. 165-IV of 4 July 2009, “On Support of the Use of Renewable Energy Sources,” as amended. Adilet Legal Information System. https://adilet.zan.kz/eng/docs/Z090000165_

5. Republic of Kazakhstan. Law No. 588-II of 9 July 2004, “On Electric Power Industry,” as amended. Adilet Legal Information System. https://adilet.zan.kz/eng/docs/Z040000588_

6. Kazakhstan Electricity Grid Operating Company. “Electricity Balance.” KEGOC Annual Report 2024. 2025. https://ar2024.kegoc.kz/en/electricity-balance.html

7. Kazakhstan Electricity Grid Operating Company. “Industry Structure.” KEGOC Annual Report 2024. 2025. https://ar2024.kegoc.kz/en/industry-structure.html

8. Kazakhstan Electricity Grid Operating Company. “Investment Activity.” KEGOC Annual Report 2024. 2025. https://ar2024.kegoc.kz/en/investment-activity.html

9. Government of the Republic of Kazakhstan. “Investment Opportunity in Renewable Energy: Kazakhstan Launches 915 MW Auction Programme for 2026.” 16 February 2026. https://www.gov.kz/memleket/entities/mfa-kuala-lumpur/press/news/details/1173394?lang=en

10. Qazaq Green. “Kazakhstan Approves 2026 Renewable Energy Auction Schedule.” 19 February 2026. https://qazaqgreen.com/en/news/kazakhstan/3386/

11. KOREM JSC. “Results of the Auction for Selection of Renewable Energy Projects,” solar auction announcement and results, 30 April 2026. https://www.korem.kz/en/news/novosti/3326

12. TotalEnergies. “Kazakhstan: TotalEnergies Signs a 25-Year PPA for a 1 GW Wind Project.” 9 June 2023. https://totalenergies.com/newsroom/kazakhstan-totalenergies-signs-25-year-ppa-1-gw-wind-project/?lang=eng

13. Kazakhstan Electricity Grid Operating Company. “Research and Development Activities.” KEGOC Annual Report 2024. 2025. https://ar2024.kegoc.kz/en/research-and-development-activities.html

14. International Energy Agency. Kazakhstan 2022: Energy Sector Review. Paris: IEA, 2022. https://www.iea.org/reports/kazakhstan-2022

15. International Energy Agency. Status of Power System Transformation 2019: Power System Flexibility. Paris: IEA, 2019. https://www.iea.org/reports/status-of-power-system-transformation-2019

16. International Energy Agency. Batteries and Secure Energy Transitions. Paris: IEA, 2024. https://www.iea.org/reports/batteries-and-secure-energy-transitions

17. International Renewable Energy Agency. Renewable Energy Auctions: Status and Trends Beyond Price. Abu Dhabi: IRENA, 2019. https://www.irena.org/publications/2019/Dec/Renewable-energy-auctions-Status-and-trends-beyond-price

18. Hirth, L. “The Market Value of Variable Renewables: The Effect of Solar Wind Power Variability on Their Relative Price.” Energy Economics 38 (2013): 218-236. https://doi.org/10.1016/j.eneco.2013.02.004

19. Ueckerdt, F., Hirth, L., Luderer, G., and Edenhofer, O. “System LCOE: What Are the Costs of Variable Renewables?” Energy 63 (2013): 61-75. https://doi.org/10.1016/j.energy.2013.10.072

20. Zerrahn, A., Schill, W.-P., and Kemfert, C. “On the Economics of Electrical Storage for Variable Renewable Energy Sources.” European Economic Review 108 (2018): 259-279. https://doi.org/10.1016/j.euroecorev.2018.07.004

21. Jurasz, J., Canales, F. A., Kies, A., Guezgouz, M., and Beluco, A. “A Review on the Complementarity of Renewable Energy Sources: Concept, Metrics, Application and Future Research Directions.” Solar Energy 195 (2020): 703-724. https://doi.org/10.1016/j.solener.2019.11.087

22. Karatayev, M., and Clarke, M. L. “A Review of Current Energy Systems and Green Energy Potential in Kazakhstan.” Renewable and Sustainable Energy Reviews 55 (2016): 491-504. https://doi.org/10.1016/j.rser.2015.10.078

23. Boute, A. “Regulatory Stability and Renewable Energy Investment: The Case of Kazakhstan.” Renewable and Sustainable Energy Reviews 121 (2020): 109673. https://doi.org/10.1016/j.rser.2019.109673

24. Mouraviev, N. “Renewable Energy in Kazakhstan: Challenges to Policy and Governance.” Energy Policy 149 (2021): 112051. https://doi.org/10.1016/j.enpol.2020.112051

25. Assembayeva, M., Egerer, J., Mendelevitch, R., and Zhakiyev, N. “A Spatial Electricity Market Model for the Power System: The Kazakhstan Case Study.” Energy 149 (2018): 762-778. https://doi.org/10.1016/j.energy.2018.02.011

26. Zhakiyev, N., Khamzina, A., Zhakiyeva, S., De Miglio, R., Bakdolotov, A., and Cosmi, C. “Optimization Modelling of the Decarbonization Scenario of the Total Energy System of Kazakhstan until 2060.” Energies 16, no. 13 (2023): 5142. https://doi.org/10.3390/en16135142

UDC 336.71:658.15(574)

Chen Jun
DBA candidate
Al-Farabi Kazakh National University
(Almaty, Kazakhstan)

CONSTRUCTION-PERIOD CREDIT RISK MANAGEMENT FOR LARGE INFRASTRUCTURE PROJECT FINANCE IN KAZAKHSTAN

Abstract. Large infrastructure projects create public value after construction has converted financing, permits, designs, equipment and civil works into an operating asset. For a bank, this conversion period concentrates credit risk because operating cash flow is limited, exposure rises with drawdown, and several events can delay completion or increase debt. This study examines the transmission channels and directional implications of construction risk for probability of default, loss given default, exposure at default, expected credit loss and prudential capital in Kazakhstan. It combines document analysis, project finance theory, lifecycle process mapping and a fully synthetic three-scenario case; it does not estimate those risk parameters or capital amounts. The proposed framework covers engineering, procurement and construction performance, cost overrun, delay, foreign-exchange mismatch, permits, offtaker obligations, environmental and social performance, drawdown control, independent review, contingency, debt service reserves, sponsor completion support, direct agreements and covenants. The analysis indicates that contractual allocation, evidence-based disbursement and forward-looking monitoring should operate together. The framework links each material risk to a measurable indicator, decision right, funding source and remedial action, providing a practical basis for banks financing transport, energy, water, municipal and industrial infrastructure in Kazakhstan.

Keywords: project finance; construction risk; credit risk; infrastructure; Kazakhstan; drawdown control; DSCR

1. Introduction

Project finance places primary reliance on project assets, contracts and future cash flows. This structure supports large investment with limited shareholder recourse, while making timely, on-budget completion decisive. Debt is drawn before stable revenue exists, progress is difficult for a bank to observe directly, and delay can increase interest, defer revenue, consume contingency and weaken counterparties.

Kazakhstan requires sustained investment in electricity, transport, water, municipal and industrial infrastructure. Its framework includes legislation on project finance, PPPs, environmental assessment and currency regulation [1-4]. The EBRD country strategy also prioritizes connectivity, green infrastructure, local-currency finance and stronger PPP frameworks [5]. These opportunities require disciplined credit processes for long construction cycles.

The financial regulator's regular Asset Quality Review, conducted in 2025 using data as of 1 January 2025 and published in 2026, provides an important banking context. The review covered 11 large banks representing 86 percent of banking-system assets and examined KZT 32.8 trillion of exposures, including KZT 3.6 trillion classified as investment projects. The aggregate capital adequacy ratio for participating banks was 17.7 percent. Supervisory adjustments raised measured Stage 2 and Stage 3 shares relative to banks' submitted

classifications [6]. This reclassification should not be interpreted as year-on-year deterioration because the same release reported improved portfolio quality compared with the 2024 review.

International rules reinforce the distinction between construction and operation. Under the Basel standardized approach, unrated project-finance exposures receive a 130 percent risk weight in the pre-operational phase and generally 100 percent in the operational phase; qualifying high-quality operational exposures may receive an 80 percent risk weight [7]. These international benchmarks should be distinguished from Kazakhstan's applicable domestic prudential rules. IFRS 9 requires lifetime expected credit losses after a significant increase in credit risk and requires forward-looking information [8]. A loan can remain contractually current while lifetime default risk has increased through delay, cost pressure, permit failure or weaker counterparties.

This article addresses three questions. First, through which channels do construction events become bank credit losses? Second, which contractual, technical and financial controls can interrupt those channels? Third, how can lenders combine scenario analysis with drawdown decisions, covenant design and IFRS 9 monitoring? The main contribution is a construction-stage control framework designed for practical use in Kazakhstan. It connects risk allocation, evidence, thresholds and remedial rights within a single lifecycle process.

2. Institutional Context and Problem Definition

2.1 The construction period as a credit discontinuity

The economics of project finance change sharply at completion. Before completion, value is represented by work in progress, equipment, land rights, permits, contracts and claims against contractors. These elements may have limited recovery value outside the original project. After completion, the asset may generate contracted or regulated revenue and can be valued through operating cash flow. Sorge describes this pattern as relatively high credit risk at project inception that tends to decline over the project life [9]. The bank's central task is therefore to control the conversion from an incomplete asset into an operating cash-flow system.

Construction risk is multidimensional. Design defects, procurement disruption or contractor default can create rework, replacement cost and delay. Currency depreciation can enlarge imported cost or foreign-currency debt. Permit, land, environmental, social or offtaker failures can interrupt critical work or acceptance. These events affect credit through common variables: completion date, cost to complete, committed funding, technical output, operating revenue, debt and recovery value.

Historical research supports conservative treatment. Cost escalation is frequent in large transport projects [10]. PPP research favors allocation to the party able to control or absorb each risk [11]. Project finance therefore relies on connected contracts, concentrated monitoring and restrictions on discretionary cash use [12-16].

2.2 Kazakhstan-specific considerations

Kazakhstan's project finance and PPP laws provide a basis for relevant transaction structures [1-2]. Environmental and currency laws affect assessment, permits and cross-currency arrangements [3-4]. Lenders must translate these frameworks into a project-specific permit matrix, security package, account structure and conditions precedent.

Foreign exchange is particularly important. Kazakhstan infrastructure projects with KZT-denominated revenue may nevertheless incur imported-equipment costs or foreign-currency debt, creating potential currency mismatch. EBRD experience shows the importance of local-currency financing for Kazakhstan infrastructure with KZT revenue [17]. A prudent structure

may combine currency-matched debt, indexed revenue, hedging, sponsor support and a limited foreign-currency tranche aligned with foreign-currency cash flow.

Public evidence also confirms the role of implementation supervision. In its 2025 board report for the Aktobe to Ulgaisyn road project, the EBRD identified project implementation risk and described engineering supervision as a principal mitigant [18]. This public-source example is used solely as institutional evidence and does not imply the author's participation in the transaction. The independent engineer does not replace the borrower's obligations, yet it gives lenders a technically qualified and independent basis for drawdown, progress and completion decisions.

3. Literature and Theoretical Foundation

Project finance can be understood through risk allocation, agency theory and information asymmetry. Participants possess different information and incentives, while the special-purpose vehicle has little history and few unencumbered assets. Contracts and monitoring make performance observable and limit opportunistic behavior. Research identifies the importance of separate project companies, non-recourse allocation, loan structure, leverage, concentrated ownership and managerial incentives [13-15].

Risk allocation is effective when the responsible party can influence the event, has capacity to perform, and retains incentives under stress. A fixed-price, date-certain EPC contract reduces uncertainty only to the extent supported by contractor capability, liability caps, security, insurance and enforceability. Sponsor completion support can close residual gaps when it is committed and the sponsor remains creditworthy.

Lender monitoring addresses the remaining information problem. Arranger certification can create economic value [19]. Independent review, controlled accounts and milestone-based drawdown test the consistency of physical progress, incurred cost and remaining funding. Direct agreements provide default notice, cure periods and possible step-in, while development-bank participation can mitigate political and coordination risk [20].

Credit pricing, accounting loss and capital are related to this architecture. Banks may price material default drivers and control other risks through project design and contracts [21]. Construction events can increase cost, delay the commercial operation date, enlarge drawn exposure and weaken forecast CFADS. These effects can raise probability of default, reduce recovery coverage and increase exposure at default. IFRS 9 expected credit loss may consequently increase, while the prudential capital effect depends on the applicable standardized or internal-ratings-based approach and its national implementation [7-8].

An effective control acts on a specific link in this transmission process. A performance bond improves potential recovery. Contingency and sponsor support address a funding gap. A completion long-stop date limits open-ended exposure. A debt service reserve account provides time for remediation after commercial operation. Scenario analysis measures the combined effect of correlated events, which is essential because delay, interest and foreign-exchange shocks often occur together.

4. Research Method

This study uses a structured qualitative and analytical design. The documentary sample comprises Kazakhstan primary legislation and supervisory releases, international prudential, accounting and environmental standards, and peer-reviewed project-finance studies available through 10 August 2026. Sources were included when they directly informed construction-risk identification, allocation, monitoring, or accounting and capital treatment. The study maps the construction lifecycle from first drawdown through provisional acceptance, performance testing

and financial completion; develops a risk-control matrix linking each risk source to evidence, decision rights, thresholds and remedial actions; and applies scenario analysis to an illustrative project to show how construction stress affects debt balance, debt service and DSCR.

The approach is conceptual, illustrative and practice-oriented. It does not estimate a statistical default model and does not use confidential bank or borrower data. No probabilities are assigned to the scenarios; the calculations test internal consistency and sensitivity rather than predict project default. DSCR is calculated as:

$DSCR_t = CFADS_t$ / Scheduled Debt Service$_t$

where CFADS is cash flow available for debt service. The analysis treats completion as both a technical and credit event. Technical completion requires the asset to satisfy agreed tests. Credit completion also requires permits, insurance, reserve funding, no continuing material default, and a sustainable operating forecast.

5. Construction-Stage Credit Risk Management Framework

5.1 Risk identification and allocation

The first layer is a pre-closing bankability review. The risk register should identify the responsible party, contractual remedy, available security and residual exposure. Table 1 presents a minimum structure.

Table 1. Construction-period credit risk transmission and control matrix

Risk source	Credit transmission	Leading evidence	Primary control	Residual risk owner
EPC design or performance failure	Rework, delay, lower output, weaker CFADS	Design review comments; unresolved interface register; failed factory or site tests	Single-point EPC responsibility where feasible; performance specifications; liquidated damages; parent guarantee; performance security	EPC contractor up to enforceable limits, then sponsor and project company
Cost overrun	Funding gap, higher leverage, delayed completion	Forecast cost to complete; contingency usage; approved and pending change orders	Fixed or capped price; committed contingency; equity funding sequence; sponsor completion support	Sponsor for agreed support, then lenders through restructuring risk
Schedule delay	More interest during construction, deferred revenue, permit expiry	Critical-path variance; float consumption; procurement delivery status	Date-certain obligation; delay damages; long-stop date; recovery plan	Contractor within caps, then sponsor and project company
Foreign exchange mismatch	Higher imported-capex cost or debt burden	Open FX position; hedge coverage; currency of remaining procurement	Currency matching; indexation; hedging; FX reserve; sponsor top-up	Project company and sponsor after hedge protection

Risk source	Credit transmission	Leading evidence	Primary control	Residual risk owner
Permit, land or connection failure	Work stoppage or inability to achieve COD	Permit register; conditions and expiry dates; land access; grid or utility milestones	Conditions precedent; legal due diligence; government or offtaker covenants; long-stop date	Party controlling the permit or connection, subject to enforceability
Offtaker or public counterparty failure	Delayed acceptance or unavailable revenue	Connection works; payment security; acceptance protocol; budget authorization	Bankable offtake agreement; direct agreement; payment support; deemed completion or deemed energy where applicable	Offtaker within contract, then political and project residual risk
Environmental and social breach	Suspension, remediation, reputational loss, lender default	ESAP status; incident reports; grievances; labor and community metrics	ESMS; independent E&S monitoring; action plan; insurance; cure and suspension rights	Project company, contractor and sponsor according to responsibility
Contractor insolvency	Work interruption, replacement premium, claims	Liquidity, arrears, subcontractor complaints, declining site resources	Financial due diligence; bonds; retention; parent guarantee; replacement rights	Contractor security first, then sponsor completion support

Source: Author's synthesis. The matrix is conceptual and requires transaction-specific legal and technical validation.

The EPC contract must be reconciled with the model, schedule, insurance, offtake agreement and loan documents. Price, schedule, completion tests, damages, security and liability caps should work as one allocation system. Environmental and social (E&S) controls also belong in the credit framework. The Equator Principles, IFC Performance Standards and EBRD policy support proportionate assessment, action plans, management systems, covenants and independent monitoring [22-24].

5.2 Conditions precedent and drawdown control

The second layer is evidence-based disbursement. Each draw should depend on eligible expenditure, verified progress, sufficient funding to complete and continuing compliance. The independent engineer confirms technical facts, while the bank retains the credit decision.

Table 2. Milestone-based drawdown control protocol

Decision point	Minimum evidence	Quantitative test	Decision and escalation
First drawdown	Executed finance and security documents; key permits; effective EPC and offtake contracts; insurance; equity evidence; baseline model and schedule	Total sources equal total uses, required equity funded or irrevocably committed, contingency meets approved minimum	Release only after all critical conditions are satisfied or formally waived by authorized credit level
Routine monthly draw	Borrower certificate; invoices; independent engineer report; site progress; account statements; updated risk register	Cumulative paid cost consistent with verified progress; no unfunded portion of the forecast cost to complete; no material covenant breach	Fund eligible amount; hold disputed or unsupported items
Major equipment payment	Factory acceptance evidence; shipping documents; title and insurance confirmation; sanctions and supplier checks	Advance-payment and performance security remain valid; exposure within approved procurement budget	Pay through controlled account or directly to supplier where documents permit
Change order	Technical justification; price and schedule impact; contractor quotation; independent engineer opinion; revised permits where needed	Remaining contingency and committed sources cover the full impact; revised DSCR meets policy	Approve within delegated threshold; escalate material variations to credit committee
Contingency use	Cause analysis; recovery plan; revised cost-to-complete forecast	Post-use contingency remains above threshold or sponsor provides replacement funding	Permit use for eligible causes; block discretionary scope additions

Decision point	Minimum evidence	Quantitative test	Decision and escalation
Provisional completion and final draw	Completion certificate; performance tests; punch list; operating permits; insurance conversion; E&S status; account funding	Output and reliability tests passed; DSRA funded; residual cost and claims fully covered	Convert to term phase only when technical and credit completion tests are met

Source: Developed by the author. The protocol is illustrative and does not replace facility-specific approval requirements.

Physical and financial progress should be reported separately because equipment payments can precede installation. The engineer should state physical completion, payment progress, forecast cost to complete and forecast completion date. The bank should reconcile invoices with controlled accounts and verify equity. Direct agreements should provide termination notice, cure and step-in periods. Security and legal opinions should cover enforceability, priority and required consent.

5.3 Forward-looking monitoring and early warning

The third layer converts monthly project information into credit indicators. Core variables are schedule and cost variance, contingency, funding sufficiency, contractor condition, permits, environmental and social actions, hedge coverage and forecast DSCR. Traffic-light thresholds require calibration to the project.

Table 3. Illustrative early-warning thresholds for lender monitoring

Indicator	Green	Amber	Red	Required lender response
Forecast delay to COD	Up to 30 days and within available float	31 to 90 days or critical float largely consumed	More than 90 days or forecast to breach the contractual long-stop date	Independent engineer recovery review; revised model; draw restriction if no credible cure
Forecast cost overrun as percent of approved budget	Below 5% and fully funded	5% to below 10%, with replacement funding documented or irrevocably committed	10% or more, or any material unfunded amount	Sponsor funding notice; contingency lock; credit committee review
Contingency consumed	Below 50% with progress proportionate	50% to 80% or consumption faster than progress	Above 80% before late-stage completion	Rebase cost to complete; prohibit discretionary changes; require replenishment
Forecast minimum DSCR after COD	At least 1.30x	1.15x to below 1.30x	Below 1.15x	Distribution lock; cash sweep; restructuring

				analysis; additional support
Contractor financial and site condition	Current payments and resources consistent with plan	Delayed subcontractor payments or declining resources	Default, abandonment, insolvency signal or prolonged work stoppage	Exercise security and replacement planning; suspend affected draws
Critical permits and connection milestones	Valid and on schedule	Renewal or approval at risk but recoverable	Missing critical authorization or connection delay threatening COD	Legal escalation; counterparty cure; revised long-stop and funding analysis
E&S action plan	Actions current; no material incident	Overdue action below 30 days or repeated minor issue	Material breach, serious incident, stop-work order or action overdue 30 days or more	Independent E&S review; corrective plan; draw suspension where justified
FX hedge or natural coverage	At least approved policy level	Temporary shortfall with executable cure	Material open position affecting funding or DSCR	Execute hedge, resize foreign-currency draw or require sponsor top-up

Source: Developed by the author. Thresholds are illustrative, not prescriptive, and should be calibrated to the transaction.

Table 3 uses illustrative thresholds. Each red indicator requires a documented decision. Correlated risks matter because a moderate delay can become severe when combined with a weak contractor, exhausted contingency and an open currency position.

Construction indicators should feed the IFRS 9 significant-increase-in-credit-risk process. Relevant evidence includes adverse cost-to-complete change, forecast covenant breach, sponsor weakening, contractor replacement, missed milestones, support restructuring and lower recovery. Banks should compare lifetime risk with origination risk, use forward-looking scenarios and retain an evidence trail [8]. Figure 1 summarizes the principal transmission channels and corresponding lender controls.

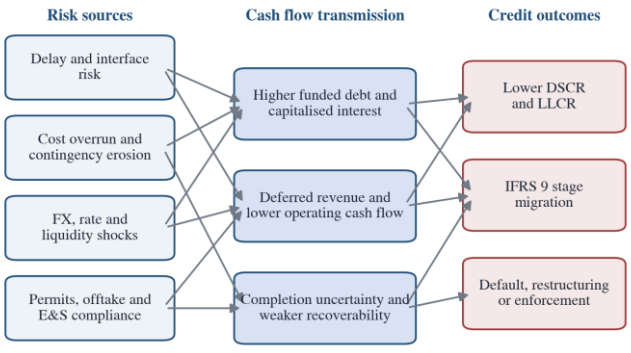


Figure 1. Transmission of construction risks into credit outcomes

Source: Developed by the author.

5.4 Completion support, reserves and covenant design

The fourth layer supplies resources and rights after an adverse event. Contingency should be committed, legally available, risk-based and restricted to eligible cost. Documentation should set the sequence among contingency, insurance, contractor recovery, equity and sponsor support. Completion support should cover defined overruns, delay interest, remediation and reserves, with payment timing and an objective release test. Sponsor capacity should be stressed for correlation with the project. The debt service reserve account (DSRA) protects early operations and must be funded at or before conversion to the term phase in accordance with the agreed funding mechanics. Covenants should address DSCR, distributions, cash sweep, additional debt, related-party payments, insurance, hedging and mandatory prepayment.

6. Illustrative Scenario Analysis

6.1 Case assumptions

The case concerns a hypothetical USD 420 million infrastructure project in Kazakhstan. It has a 36-month construction period, 70 percent initial debt financing and a 15-year repayment period after completion. The asset, sector, sponsors, counterparties and contracts are intentionally unspecified. Every number in this section is illustrative.

The base scenario assumes USD 294 million of debt, a 7.5 percent effective debt rate, annual CFADS of USD 52 million and annual debt service of USD 33.31 million. The moderate case assumes that delay, interest and debt-funded construction pressure raise debt at COD to USD 335 million and the rate to 8.5 percent, while CFADS declines to USD 46.8 million. The severe case assumes debt of USD 376.8 million, a 9.5 percent rate and CFADS of USD 41.6 million. Annual debt service is calculated as a level annual payment, $ADS = D \times r / [1 - (1 + r)^{-15}]$, using fixed annual rates and year-end payments. The moderate and severe scenarios correspond to COD at months 42 and 48, debt-at-COD increases of 13.9 percent and 28.2 percent, interest-rate increases of 100 and 200 basis points, and CFADS reductions of 10 percent and 20 percent relative to the base case. The stressed debt balances summarize assumed capitalized interest and funded completion costs; they do not imply that lenders are obliged to finance overruns without separate approval.

Figure 2 illustrates the base, moderate and severe completion paths used in the scenario analysis.

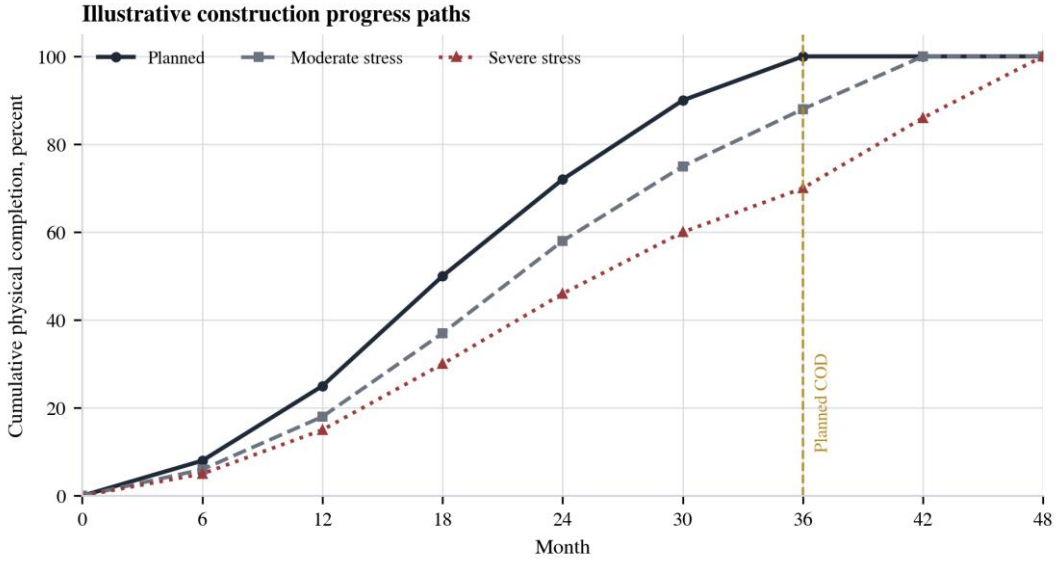


Figure 2. Illustrative cumulative construction progress under three scenarios

Source: Author's illustrative construction. All values are hypothetical.

Table 4. Illustrative construction stress and post-completion debt capacity

Metric	Base case	Moderate case	Severe case
Initial project cost, USD million	420.0	420.0	420.0
COD month	36	42	48
Delay from plan	0	6 months	12 months
Debt at COD, USD million	294.0	335.0	376.8
Increase in debt at COD from base	0.0%	13.9%	28.2%
Debt as percent of initial project cost	70.0%	79.8%	89.7%
Effective debt rate	7.5%	8.5%	9.5%
Repayment period	15 years	15 years	15 years
Annual CFADS, USD million	52.0	46.8	41.6
CFADS change from base	0.0%	-10.0%	-20.0%
Annual debt service, USD million	33.31	40.34	48.13
DSCR	1.56x	1.16x	0.86x
Indicative credit interpretation	Resilient headroom	Thin headroom and likely restrictions	Debt service not covered without restructuring or support

Source: Author's fully synthetic calculations. Values are illustrative and not prescriptive.

The moderate case illustrates a common credit transition. Construction may still reach completion, yet additional debt and lower operating cash flow reduce DSCR from 1.56x to 1.16x. This outcome can activate a distribution lock, cash sweep, additional reserve requirement or sponsor support before any scheduled payment is missed. The severe case produces DSCR of 0.86x. Annual CFADS is USD 6.53 million below scheduled debt service, so the project cannot sustain the original debt profile from operations alone.

The scenario also shows why a lender should monitor debt at completion, not only percentage cost overrun. Delay increases interest and fees; currency movements can increase imported cost; and deferred revenue weakens liquidity. A project may complete physically while failing its credit completion test. Possible responses include additional equity, extension of tenor, temporary principal grace, a formally approved margin restructuring where justified by sustainable repayment capacity, debt reduction, stronger offtaker support, or a combination. Any modification requires renewed valuation, legal review, IFRS 9 assessment and authorization under the bank's restructuring policy.

Figure 3 compares debt burden, operating cash flow and coverage outcomes across the three scenarios.

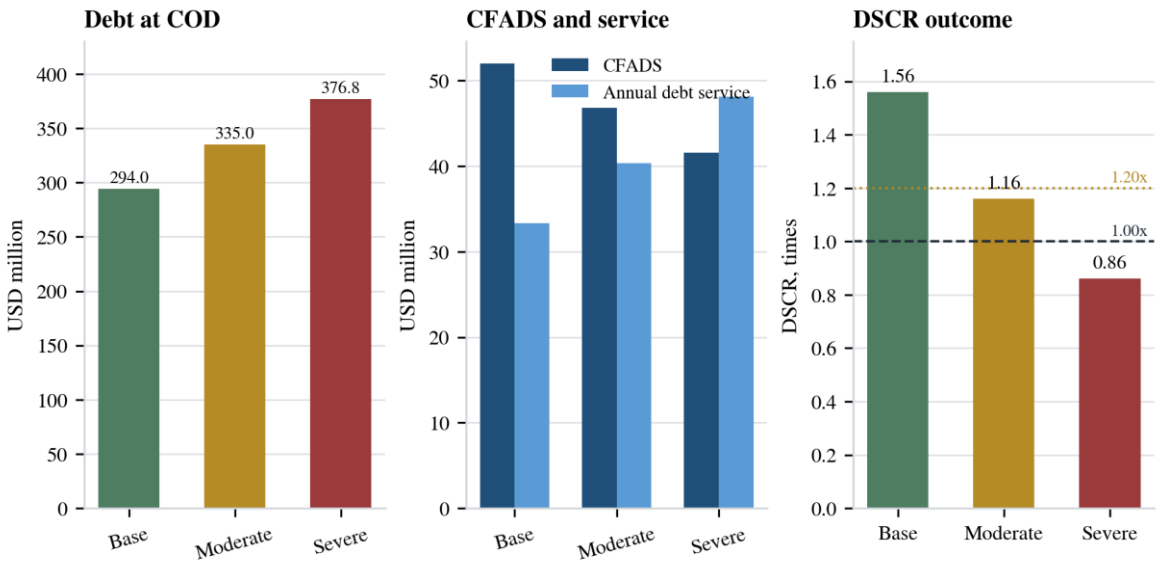


Figure 3. Illustrative debt burden, cash flow and coverage outcomes

Source: Author's calculation using the illustrative case assumptions.

7. Policy and Practice Recommendations

7.1 Recommendations for lending banks

Banks should establish a specialized construction-credit function integrating relationship, credit, engineering, legal, E&S, insurance, model and operations expertise. Independent advisers provide evidence, while internal officers retain responsibility for default and recovery analysis.

The approved model should be a controlled credit document. Material change orders, delays and funding amendments should update cost, schedule, debt at COD, CFADS, coverage ratios and covenant headroom. Models must reconcile to loan accounts and engineer forecasts. Routine certified draws may use delegated authority, while material change, contingency depletion, sponsor weakening or forecast breach should return to the credit committee. Waivers should state the basis, duration, compensating controls and follow-up date.

Construction indicators should enter IFRS 9 and watchlist systems because technical delay, permit failure and cost-to-complete deficiency may precede payment arrears. The credit file should link scenarios to staging, PD, LGD and ECL, consistent with the supervisory emphasis demonstrated by the AQR [6]. Pricing, economic capital and concentration limits should reflect construction duration, contractor concentration, currency mismatch, public-counterparty dependence and completion support, alongside the Basel lifecycle distinction [7].

7.2 Recommendations for transaction documentation

Loan documents should define completion through objective technical, legal, E&S and financial tests. Core provisions include a long-stop date, overrun funding, sponsor completion obligations, controlled accounts, information and inspection rights, drawstops, prepayment, reserves and distribution limits. Direct and intercreditor agreements should coordinate notice, cure, step-in, voting, enforcement and recoveries. EPC damages, liability caps, security and insurance should be tested against debt capacity. Where contracts are split, the project company needs one integrated schedule and interface matrix.

7.3 Recommendations for Kazakhstan's financing ecosystem

Standardized direct agreements, permit timelines, payment mechanisms and security registration can improve bankability. KZT revenue supports local-currency finance, subject to

rate and refinancing stress [17]. Syndication shares exposure and expertise; common information and development-bank participation may reduce political and coordination risk [20].

8. Limitations and Future Research

This study has four limitations. It uses an illustrative case rather than loan-level Kazakhstan data. Its thresholds are examples and require calibration by sector, technology, contract and bank risk appetite. It addresses the lender's perspective and gives less attention to sponsor and public-authority optimization. It also treats major risks within a unified framework without estimating their empirical correlations.

Future research could build a Kazakhstan project finance database containing original budget, final cost, planned and actual COD, financing currency, contractor structure, covenant outcomes and recovery experience. Econometric analysis could then test which construction indicators best predict Stage 2 transfer, restructuring and default. Comparative research could also examine local-currency and foreign-currency structures, development-bank participation and differences among energy, transport, water and industrial projects.

9. Conclusion

Construction-period credit risk is the central conversion risk of project finance. During this stage, lenders fund an asset that has yet to demonstrate operating capacity, while cost, schedule, currency, permits, counterparties and environmental or social performance can change simultaneously. Kazakhstan's infrastructure needs and project-finance framework create substantial lending opportunities, accompanied by a requirement for disciplined lifecycle control.

The analysis shows that robust credit management requires an integrated system. Contracts allocate controllable risks. Independent review makes progress observable. Drawdown conditions prevent unsupported exposure growth. Contingency, insurance and sponsor completion support provide resources for identified gaps. Direct agreements and covenants create decision rights. Scenario analysis translates technical events into debt, CFADS and DSCR outcomes, with directional implications for expected credit loss and approach-dependent prudential capital. The illustrative case demonstrates how a project can move from 1.56x DSCR to 1.16x or 0.86x when construction pressure increases debt and reduces cash flow.

The practical principle is straightforward: each material construction risk should have an identified owner, measurable indicator, committed funding source and enforceable lender response. When these four elements are connected and updated throughout construction, a bank can support infrastructure development while maintaining prudent credit quality and capital discipline.

References:

1. Republic of Kazakhstan. (2006). Law No. 126, On Project Financing and Securitization. Adilet Legal Information System. https://adilet.zan.kz/eng/docs/Z060000126_

2. Republic of Kazakhstan. (2015). Law No. 379-V, On Public-Private Partnership. Adilet Legal Information System. https://adilet.zan.kz/eng/docs/Z1500000379

3. Republic of Kazakhstan. (2021). Environmental Code of the Republic of Kazakhstan, Code No. 400-VI. Adilet Legal Information System. https://adilet.zan.kz/eng/docs/K2100000400

4. Republic of Kazakhstan. (2018). Law No. 167-VI, On Currency Regulation and Currency Control. Adilet Legal Information System. https://adilet.zan.kz/eng/docs/Z1800000167

5. European Bank for Reconstruction and Development. (2022). Strategy for Kazakhstan 2022 to 2027. https://www.ebrd.com/home/what-we-do/where-we-invest/kazakhstan.html

6. Agency of the Republic of Kazakhstan for Regulation and Development of the Financial Market. (2026). The Agency Presented the Results of the Regular AQR for 2025. Official release. https://www.gov.kz/memleket/entities/ardfm/press/news/details/1152793?lang=ru

7. Basel Committee on Banking Supervision. (2025). Basel Framework, CRE20: Standardised Approach, Individual Exposures. Bank for International Settlements. https://www.bis.org/basel_framework/chapter/CRE/20.htm

8. IFRS Foundation. (2025). IFRS 9 Financial Instruments. https://www.ifrs.org/content/dam/ifrs/publications/html-standards/english/2025/issued/ifrs9.html

9. Sorge, M. (2004). The nature of credit risk in project finance. BIS Quarterly Review, December, 91-101. https://www.bis.org/publ/qtrpdf/r_qt0412h.pdf

10. Flyvbjerg, B., Holm, M. K. S., & Buhl, S. L. (2003). How common and how large are cost overruns in transport infrastructure projects? Transport Reviews, 23(1), 71-88. https://doi.org/10.1080/01441640309904

11. Grimsey, D., & Lewis, M. K. (2002). Evaluating the risks of public-private partnerships for infrastructure projects. International Journal of Project Management, 20(2), 107-118. https://doi.org/10.1016/S0263-7863(00)00040-5

12. World Bank Group, Asian Development Bank, and Inter-American Development Bank. (2017). Public-Private Partnerships Reference Guide, Version 3.0. https://openknowledge.worldbank.org/entities/publication/e64a94b9-b0cb-51a3-b8dd-36bd7944284f

13. Brealey, R. A., Cooper, I. A., & Habib, M. A. (1996). Using project finance to fund infrastructure investments. Journal of Applied Corporate Finance, 9(3), 25-39. https://doi.org/10.1111/j.1745-6622.1996.tb00296.x

14. Kleimeier, S., & Megginson, W. L. (2000). Are project finance loans different from other syndicated credits? Journal of Applied Corporate Finance, 13(1), 75-87. https://doi.org/10.1111/j.1745-6622.2000.tb00043.x

15. Esty, B. C. (2004). Why study large projects? An introduction to research on project finance. European Financial Management, 10(2), 213-224. https://doi.org/10.1111/j.1354-7798.2004.00247.x

16. Yescombe, E. R. (2014). Principles of Project Finance (2nd ed.). Academic Press.

17. European Bank for Reconstruction and Development, Independent Evaluation Department. (2025). Evaluation of EBRD Local Currency Financing: Country Case Studies. https://www.ebrd.com/content/dam/ebrd_dxp/assets/pdfs/evaluation/2025/local-currency-operations-country-case-studies.pdf

18. European Bank for Reconstruction and Development. (2025). Kazakhstan Aktobe to Ulgaisyn Road: Board Report.

https://www.ebrd.com/content/dam/ebrd_dxp/documents/project/55068/Kazakhstan-Aktobe-Ulgaisyn%20Road%20%28redacted_final%29.pdf

19. Gatti, S., Kleimeier, S., Megginson, W. L., & Steffanoni, A. (2013). Arranger certification in project finance. Financial Management, 42(1), 1-40. https://doi.org/10.1111/j.1755-053X.2012.01210.x

20. Hainz, C., & Kleimeier, S. (2012). Political risk, project finance, and the participation of development banks in syndicated lending. Journal of Financial Intermediation, 21(2), 287-314. https://doi.org/10.1016/j.jfi.2011.10.002

21. Blanc-Brude, F., & Strange, R. (2007). How banks price loans to public-private partnerships: Evidence from the European markets. Journal of Applied Corporate Finance, 19(4), 94-106. https://doi.org/10.1111/j.1745-6622.2007.00163.x

22. Equator Principles Association. (2020). The Equator Principles, EP4. https://equator-principles.com/app/uploads/The-Equator-Principles_EP4_July2020.pdf

23. International Finance Corporation. (2012). Performance Standards on Environmental and Social Sustainability. https://www.ifc.org/content/dam/ifc/doc/2010/2012-ifc-performance-standards-en.pdf

24. European Bank for Reconstruction and Development. (2024). Environmental and Social Policy 2024. https://www.ebrd.com/home/news-and-events/publications/institutional-documents/environmental-and-social-policy-2024.html

UDC 629.331

Alniyaz Danial
student,
Shoqan Ualikhanov Private School
Scientific supervisor: Bitegenova Aliya Armankyzy
(Almaty, Kazakhstan)

HOW MUCH FUEL DOES A HYBRID REALLY SAVE? A SAME-MODEL COMPARISON OF HYBRID AND GASOLINE CARS IN US FUEL ECONOMY DATA

Abstract: Hybrid cars are usually advertised as economical, but the saving depends on the car and on how it is driven. This study compares hybrid and conventional gasoline versions of the same model in the official US fuel economy database of the Environmental Protection Agency and the Department of Energy. A hybrid was paired with a conventional car when both had the same make, base model, model year and drive type, which gave 497 pairs covering about 60 models from 23 makes and model years 2003 to 2027. Before seeing any results, I predicted that hybrids save more fuel in the city than on the highway, and that mild hybrids with small batteries and heavy, powerful vehicles that are driven mostly on the highway save almost nothing. For the median pair, a hybrid was rated to use 36% less fuel in the city, 14% less on the highway and 28% less in combined driving. The city saving was larger than the highway saving in 494 of 497 pairs and for all 60 models, and the result held when only versions with the same engine were compared. Twenty-one pairs saved less than 5% in combined driving. All of them were SUVs or pickup trucks, and twenty carried the database label "Mild Hybrid". In the nine pairs where the hybrid was rated to use more fuel, the hybrid version also had a larger engine than every conventional version, although one mild hybrid Jeep Wrangler was also worse than a conventional Wrangler with the same engine. Pickup trucks saved little even without the mild hybrid label. These are official test ratings, not measurements in real traffic, but they suggest that the size and purpose of the vehicle matter at least as much for fuel saving as the type of hybrid system.

Keywords: hybrid electric vehicles; fuel economy; mild hybrids; city and highway driving; EPA fuel economy data

1. Introduction

I am interested in mechanical engineering and engines, and I chose this topic because hybrids are now one of the most common ways car makers reduce fuel consumption. A hybrid combines a gasoline engine with an electric motor and a battery. It can save fuel because the electric system can restart the engine after a stop, recover energy during braking, help the engine during acceleration and, in some hybrids, drive the car on electric power alone. Not all hybrids can do all of this. Hybrids are usually classified by their degree of hybridization into micro, mild, medium and full hybrids: a mild hybrid offers engine start-stop and regenerative braking, while medium and full hybrids can also drive the car with the electric motor [1]. A simulation of a 48-volt mild hybrid system added to a passenger car found that it reduced fuel consumption by 7.7% on the NEDC test cycle and by 1.7% on a real driving cycle [2].

The saving also depends on how the car is driven. In tests of one conventional gasoline car and its hybrid version, Wang and co-authors found that the hybrid's fuel consumption on the

Chinese CLTC test cycle was 20.23% lower than on the WLTC cycle, while for the conventional car it was 8.41% higher, and concluded that vehicle application scenarios must be considered to obtain better fuel efficiency [3]. Many hybrids also use engines that work on the Atkinson cycle instead of the usual Otto cycle [4].

The official US fuel economy database contains many models sold both as a hybrid and as a conventional car in the same year. This makes it possible to compare a hybrid with the same car without the hybrid system, which removes most differences in size, body and brand. This study asks three questions. How much fuel does a hybrid save compared with the same model on gasoline? Is the saving larger in the city than on the highway? And which hybrids save almost nothing?

2. Data and methods

2.1. Data

All data come from the vehicle file of the fueleconomy.gov website, administered by Oak Ridge National Laboratory for the US Department of Energy and the Environmental Protection Agency (EPA), which contains fuel economy information for model years 1984 to the present [5]. The file was downloaded on 29 September 2026 and contains 50,407 vehicles. For each vehicle, the database gives the EPA city, highway and combined fuel economy in miles per gallon (fields city08, highway08 and comb08), the engine displacement and description, the drive type and the EPA size class [5]. The city rating comes from a test that represents stop-and-go rush hour traffic with 23 stops and an average speed of 21.2 mph, and the highway rating from a test of free-flowing traffic with no stops and an average speed of 48.3 mph [6]. The database marks 1,899 vehicles as hybrids. A conventional vehicle was defined as one with no alternative fuel type and regular, midgrade or premium gasoline as its fuel.

2.2. Pairs

A hybrid was paired with conventional cars of the same model year, make, base model and drive type. The base model was obtained by removing the words "Hybrid" and "HEV" and common trim and drive words from the model name, for example "Camry Hybrid LE" and "Camry LE" both become "camry". When several conventional versions matched, their median value was used. One false match was removed by hand: the 2016 Chevrolet Malibu Limited, the previous generation sold alongside the new Malibu, was not used as a conventional version. This gave 497 pairs. They cover 71 model names, which correspond to about 60 models (for example, six versions of the Jeep Wrangler name), from 23 makes and model years 2003 to 2027; some models have several hybrid versions in the same year, so the pairs fall into 440 different combinations of model year, make, model and drive type. Most hybrids in the database did not find a pair, either because no conventional version was sold or because the model names differ.

2.3. Measures

The fuel saved by a hybrid was calculated as 1 − MPG(conventional) / MPG(hybrid). This is the share of fuel the hybrid saves over the same distance. For example, a conventional car rated at 30 MPG and a hybrid at 40 MPG give a saving of 1 − 30/40 = 25%. The saving was calculated separately for city, highway and combined ratings.

My predictions were recorded on 4 October 2026, before any results were calculated. The analysis plan was written on 7 October 2026 together with the first calculation; it specified the pairing rule, the three savings, the comparison of city and highway savings, and the list of pairs that save less than 5% in combined driving. On 13 October the pairing rule was corrected to remove false matches, such as the Mitsubishi Outlander paired with the Outlander Sport, which

reduced the number of pairs from 502 to 497. The following checks were also added after the first review of the paper: a comparison using one value per model, a comparison of only those versions that have the same engine (the same displacement, number of cylinders and turbocharger or supercharger), the grouping by vehicle class and by the database label "Mild Hybrid", and the Wilcoxon signed-rank test for the city and highway savings. Because the savings are not normally distributed, medians and interquartile ranges are reported.

3. Results

3.1. Saving in the city and on the highway

For the median pair, a hybrid was rated to use 35.7% less fuel than the same model on gasoline in the city, 13.5% less on the highway and 27.6% less in combined driving (Table 1). The city saving was larger than the highway saving in 494 of 497 pairs (Wilcoxon signed-rank test, $p < 0.001$). Because some models appear in many pairs, the comparison was repeated with one median value per model: the savings were 35.1% in the city, 13.5% on the highway and 28.5% combined, and the city saving was larger for all 60 models ($p < 0.001$). Figure 1 shows these 60 models: all of them lie above the line where the city and highway savings would be equal.

Some hybrid versions have a different engine from the conventional versions they were compared with. For 199 pairs (33 models) a conventional version with the same engine was available, meaning the same displacement, number of cylinders and turbocharging. Compared only with these, the hybrid saved 33.9% in the city, 9.7% on the highway and 25.0% combined, and the city saving was larger in 198 of 199 pairs. For example, the 2024 Toyota Camry hybrid saved 32-39% in combined driving compared with the conventional Camry with the same 2.5-liter engine, depending on the version. The 2024 Ford F-150 hybrid with four-wheel drive saved 27% in the city, but on the highway it saved nothing: both versions are rated at 24 MPG.

Table 1. Median fuel saved by the hybrid compared with the same model on gasoline, %

Group	Pairs	City	Highway	Combined
All pairs	497	35.7	13.5	27.6
Same engine	199	33.9	9.7	25.0
Cars and vans	204	40.0	20.3	32.9
SUVs	231	34.9	10.0	25.6
Pickup trucks	62	14.7	4.2	10.5
Labeled "Mild Hybrid"	50	11.8	4.2	10.5
Not labeled "Mild Hybrid"	447	37.0	15.6	29.3

Note: "Same engine" compares each hybrid only with conventional versions with the same displacement, number of cylinders and turbocharging. The interquartile range for all pairs is 25.0-41.4% in the city, 6.2-20.6% on the highway and 18.5-33.3% combined.

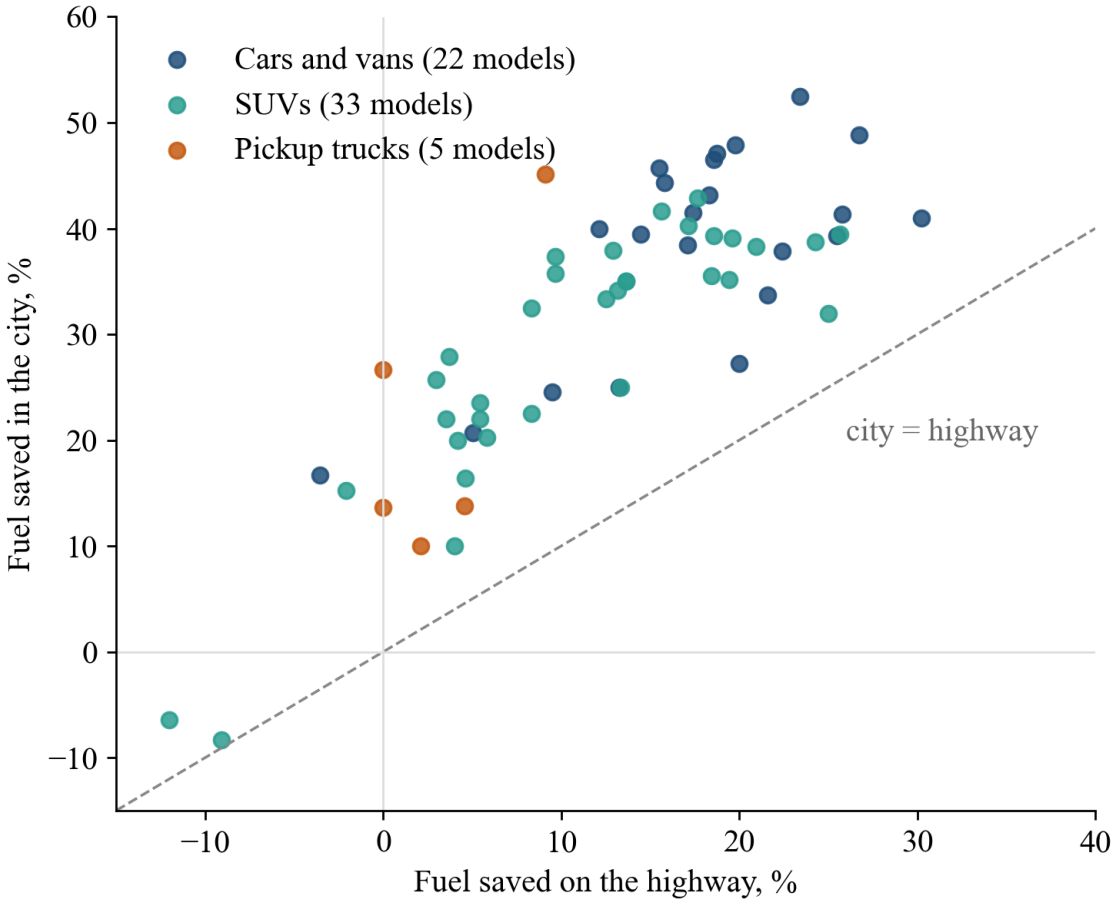


Figure 1. Median fuel saved by the hybrid in the city and on the highway for 60 models. Points above the dashed line save more in the city.

3.2. Hybrids that save almost nothing

Twenty-one pairs saved less than 5% of fuel in combined driving, and in nine of them the hybrid version was rated as using more fuel than the conventional one. All twenty-one were SUVs or pickup trucks (versions of the Audi Q7, Jeep Wrangler, Jeep Wagoneer and Ram 1500), and twenty of them carry the label "Mild Hybrid" in the engine description of the database; the remaining one, a 2027 Ram 1500 with two-wheel drive, is labeled "Hybrid". In all nine pairs where the hybrid was rated to use more fuel, the hybrid version also had a larger engine than every conventional version it was compared with: a 3.0-liter engine against 2.0 liters in the Audi Q7, and a 5.7-liter V8 against 3.0 liters in the Jeep Wagoneer and Ram 1500. Where a conventional version with the same engine existed, as for the Jeep Wrangler and the 2027 Ram 1500, the low-saving hybrids saved between −9.5% and 10.5% compared with it; the 2020 Jeep Wrangler 4WD with a 2.0-liter turbo engine was 9.5% worse than the conventional Wrangler with the same engine.

Hybrids labeled "Mild Hybrid" saved a median of 10.5% in combined driving, compared with 29.3% for the other hybrids. This label appears in the database only from model year 2018. It is carried by 841 hybrids in the database, but only 50 of them, from Audi, Jeep and Ram, could be paired. In 16 of these 50 pairs the hybrid had a smaller engine than some of the conventional versions, which raises its saving; compared only with versions with the same engine, the 20 mild hybrid pairs saved 8.8% in combined driving. Vehicle class mattered as well. Pickup trucks saved a median of 10.5%, compared with 32.9% for cars and vans, and even

pickup trucks without the mild hybrid label, such as the Toyota Tundra and Tacoma hybrids and the Ford F-150 hybrid, saved a median of only 9.1% (33 pairs).

4. Discussion

My first prediction was that hybrids save more fuel in the city. The data support it clearly, for all pairs, for each model and for versions with the same engine. I explain this by how a hybrid works. In the city the car constantly accelerates and brakes, and the EPA city test includes 23 stops [6]. During braking, regenerative braking returns energy to the battery, and at low speed a full hybrid can move on electric power alone. On the highway the speed is steady, these advantages almost disappear, and most of the work is done by the gasoline engine. This agrees with the finding that the fuel consumption of a hybrid depends strongly on the driving cycle [3].

But the highway saving is not zero: the median pair still saves about 14% on the highway, where there is little braking to recover energy from. In my opinion, the reason is the engine itself. In many hybrids the gasoline engine works on the Atkinson cycle [4], and, in my view, this cycle gives higher efficiency at constant speed than the usual Otto cycle. In addition, the electric motor takes the peak loads during overtaking or climbing, so the engine can stay in its most efficient range and does not have to enrich the mixture or run at inefficient speeds. Wang and co-authors explained the much lower NOx emissions of the hybrid they tested by lower engine load and less fuel enrichment [3], which agrees with the idea that the hybrid's engine works in a more favorable range. The data in this study do not test this explanation directly; they only show that the highway saving is positive for most pairs.

My second prediction was that mild hybrids with small batteries and heavy, powerful vehicles driven mostly on the highway save almost nothing. The data support it partly. The hybrids that saved least were all SUVs and pickup trucks, almost all of them with modern mild hybrid systems, and a simulated 48-volt mild hybrid system saved only 1.7-7.7% of fuel [2]. But two other things were also important. First, in most pairs where the hybrid was worse, it also had a much larger engine, so in these cases the hybrid version is in effect a more powerful car, not just the same car with a hybrid system; the one exception is the 2020 Jeep Wrangler with the same 2.0-liter turbo engine. Second, pickup trucks saved little even without the mild hybrid label. The size and purpose of the vehicle therefore seem to matter at least as much as the type of hybrid. My explanation was that the extra weight of the hybrid system increases fuel consumption, but the database does not contain vehicle weight, so this could not be checked. My own first look at three pairs on the fueleconomy.gov website showed the same pattern before I saw the full data: the Toyota RAV4 hybrid saved much more in the city than on the highway, and the Chevrolet Corvette E-Ray, a powerful sports car, was rated almost the same as the conventional Corvette. These quick comparisons did not follow the pairing rule of this study exactly.

For a buyer, this means that a hybrid badge alone does not guarantee a real saving. A hybrid car used mostly in the city saves the most. An SUV or pickup truck with a hybrid system, especially a mild one, may save very little.

5. Limitations

The results are based on official EPA test ratings, not on measurements in real traffic, and the ratings are rounded to whole miles per gallon, so a difference of one mile per gallon can change a small saving by several percentage points. The pairs were formed by model name, so a hybrid and a conventional version of the same model can still differ in engine, power or equipment; the same-engine comparison reduces but does not remove this problem. Only some of the hybrids in the database could be paired, so the results describe these models, not all

hybrids. The label "Mild Hybrid" appears only from model year 2018 and only 50 of the 841 labeled hybrids could be paired: older hybrids that Xue and co-authors list as examples of mild or medium hybrids, such as the Saturn Vue, Chevrolet Malibu and Honda Civic [1], carry no such label and are included in the other group. The database does not separate all-wheel and four-wheel drive in the same way for all years, which may affect the matching by drive type. Finally, the data describe cars sold in the United States, which are not the same as the cars sold in Kazakhstan.

6. Conclusion

In 497 pairs of hybrid and conventional gasoline versions of the same model, the hybrid was rated to use a median of 36% less fuel in the city, 14% less on the highway and 28% less in combined driving, and the city saving was larger for every one of the 60 models. All hybrids that saved less than 5% were SUVs and pickup trucks, almost all with mild hybrid systems, and almost all of those rated worse than the conventional version also had larger engines. A hybrid saves the most when it is a car driven in the city, and the least when it is a large, powerful vehicle with a small hybrid system or a bigger engine than the conventional version.

Data availability

All data are publicly available from fueleconomy.gov [5]. The list of pairs and the scripts that calculate every number in this paper are available from the author on request.

AI and tools disclosure

I chose the research question, wrote my predictions before seeing the results for the whole database, compared the first three pairs on the fueleconomy.gov website, found the study by Wang and co-authors, and wrote, in Russian, the explanation of why hybrids save fuel on the highway. My mentor downloaded the full database, wrote the code for the calculations with the help of generative AI tools and prepared the figure. The English text was drafted by my mentor with the help of generative AI tools, based on the results and on my own written answers, and I reviewed and approved it. No data, results or citations were generated by AI: all numbers come from the fueleconomy.gov database, and all cited sources were checked against the originals. I take full responsibility for the final manuscript.

Conflicts of interest

The author declares no conflicts of interest.

References:

1. Xue Q., Zhang X., Teng T., Zhang J., Feng Z., Lv Q. A comprehensive review on classification, energy management strategy, and control algorithm for hybrid electric vehicles // Energies. 2020. Vol. 13, No. 20. Art. 5355. DOI: 10.3390/en13205355.

2. Abidin S.F.Z., Khalid A., Zanalli S., Zahari I., Jalal R.I.A., Abas M.A., Koten H. The effect of 48V mild hybrid technology on fuel consumption of a passenger car by using simulation cycle // Case Studies in Thermal Engineering. 2021. Vol. 28. Art. 101492. DOI: 10.1016/j.csite.2021.101492.

3. Wang Y., Hao C., Ge Y., Hao L., Tan J., Wang X., Zhang P., Wang Y., Tian W., Lin Z., Li J. Fuel consumption and emission performance from light-duty conventional/hybrid-electric vehicles over different cycles and real driving tests // Fuel. 2020. Vol. 278. Art. 118340. DOI: 10.1016/j.fuel.2020.118340.

4. León R., Montaleza C., Maldonado J.L., Tostado-Véliz M., Jurado F. Hybrid electric vehicles: a review of existing configurations and thermodynamic cycles // Thermo. 2021. Vol. 1, No. 2. P. 134-150. DOI: 10.3390/thermo1020010.

5. U.S. Department of Energy, U.S. Environmental Protection Agency. fueleconomy.gov web services: vehicle data file and data description, model years 1984-2027 [Electronic resource]. URL: https://www.fueleconomy.gov/feg/ws/index.shtml (data downloaded: 29.09.2026; accessed: 13.10.2026).

6. U.S. Department of Energy, U.S. Environmental Protection Agency. Detailed test information: fuel economy test schedules [Electronic resource]. URL: https://www.fueleconomy.gov/feg/fe_test_schedules.shtml (accessed: 13.10.2026).

Central Asian Scientific Journal

ТЕХНИКАЛЫҚ ҒЫЛЫМДАР - ТЕХНИЧЕСКИЕ НАУКИ - TECHNICAL SCIENCE

UDC 621.39:681.7

Nurmakhan Almira

Bachelor Student
Scientific supervisor: Mamilov Bakhytzhan
Associate Professor
Department of Telecommunication Engineering
Muhammadjan Tynyshbayev ALT University
(Almaty, Kazakhstan)

ADVANTAGES OF FIBER OPTIC COMMUNICATION OVER TRADITIONAL COPPER CABLES

Abstract: This research paper investigates the technological, economic, and environmental advantages of fiber optic communication compared to traditional copper cable transmission systems. The study employs a mixed-method research design combining quantitative performance evaluation and qualitative expert analysis in telecommunication infrastructure projects. Key variables considered in the analysis include Data Transmission Capacity (DTC), Signal Loss Rate (SLR), Energy Efficiency Level (EEL), Installation Cost Efficiency (ICE), and Network Reliability Index (NRI). Statistical results demonstrate that fiber optic systems significantly outperform copper-based communication networks in bandwidth capacity, transmission distance, and electromagnetic interference resistance. Structural analysis further reveals that technological innovation and network modernization policies act as mediating factors in improving communication efficiency. Cluster comparisons show that regions adopting fiber optic infrastructure experience higher data speed stability, lower maintenance requirements, and enhanced digital service quality. Overall, the findings confirm that transitioning from copper-based networks to fiber optic communication provides substantial benefits for sustainable telecommunication development and smart urban infrastructure systems.

Keywords: Fiber Optic Communication, Copper Cables, Network Efficiency, Data Transmission, Telecommunication Infrastructure, Technological Innovation.

Introduction. Modern telecommunication systems require high-speed, reliable, and energy-efficient data transmission technologies. Traditional copper cable communication networks have been widely used due to their affordability and ease of installation. However, the rapid growth of internet traffic, cloud computing, and smart technologies has exposed limitations in copper-based systems such as signal attenuation, bandwidth constraints, and electromagnetic interference.

Fiber optic communication represents an advanced transmission technology that uses light signals instead of electrical signals to transfer data. This shift from electrical to optical transmission significantly improves communication performance and network sustainability. As digital transformation accelerates worldwide, understanding the comparative advantages of

fiber optics over copper cables becomes crucial for policymakers, engineers, and urban planners.

A. Understanding the Evolution of Telecommunication Technologies

The rapid development of modern information and communication technologies has significantly transformed the way data is transmitted and processed across global networks. Increasing reliance on digital platforms, cloud services, smart devices, and high-definition multimedia applications has created an unprecedented demand for faster, more reliable, and energy-efficient communication systems. In response to these growing requirements, telecommunication infrastructures are undergoing continuous technological evolution aimed at improving transmission capacity, reducing latency, and ensuring network stability.

Historically, copper cable communication systems have served as the backbone of telecommunication networks, supporting voice transmission, broadband internet services, and local area connectivity. Their widespread adoption was largely driven by cost-effectiveness, ease of installation, and compatibility with existing electrical transmission technologies. However, the exponential growth in data traffic and the emergence of advanced digital applications have revealed the technical limitations of copper-based communication systems, necessitating the exploration of more advanced transmission media.

B. Identifying Limitations of Traditional Copper Cable Systems

Despite their historical importance, traditional copper cables face several inherent disadvantages that hinder their performance in modern high-capacity communication environments. One of the primary challenges associated with copper transmission is signal attenuation, which increases with distance and frequency, resulting in reduced data quality and transmission efficiency. Additionally, copper cables are highly susceptible to electromagnetic interference and environmental disturbances, leading to signal degradation and network instability.

Bandwidth constraints also limit the ability of copper networks to support emerging technologies such as high-speed broadband services, real-time data analytics, and Internet of Things (IoT) applications. As communication networks become more complex and data-intensive, the operational costs associated with maintaining copper infrastructure—including energy consumption, frequent signal amplification, and system upgrades—continue to rise. These challenges highlight the urgent need for alternative communication technologies capable of meeting the performance demands of next-generation digital systems.

C. Exploring the Role of Fiber Optic Communication as an Advanced Solution

Fiber optic communication has emerged as a revolutionary transmission technology that utilizes light signals transmitted through thin strands of glass or plastic fibers to carry information over long distances. Unlike electrical signal transmission in copper cables, optical communication provides significantly higher bandwidth capacity, minimal signal loss, and enhanced resistance to electromagnetic interference. These characteristics enable fiber optic systems to deliver superior data transmission speeds, improved network reliability, and greater scalability in expanding communication infrastructure.

Furthermore, fiber optic networks support the integration of advanced digital technologies, including 5G communication systems, smart city platforms, artificial intelligence applications, and large-scale cloud computing services. The ability of optical fibers to transmit large volumes of data simultaneously makes them a critical component of modern telecommunication strategies aimed at enhancing global connectivity and technological innovation.

D. Emphasizing the Strategic Importance of Transitioning to Fiber Optic Infrastructure

The transition from conventional copper-based communication networks to fiber optic infrastructure represents not only a technological advancement but also a strategic investment in sustainable digital development. Fiber optic systems generally require lower energy consumption, offer longer operational lifespans, and reduce maintenance frequency compared to traditional transmission technologies. These advantages contribute to improved economic efficiency and reduced environmental impact in telecommunication operations.

Moreover, the deployment of fiber optic networks plays a crucial role in supporting national digital transformation initiatives, enabling smart urban development, and strengthening global information exchange systems. Understanding the comparative benefits of fiber optic communication is therefore essential for engineers, policymakers, and infrastructure planners seeking to design resilient and future-oriented communication networks capable of addressing the evolving needs of modern society.

This study aims to provide a comparative empirical assessment of fiber optic and copper cable communication efficiency

Methodology.

Conceptual Framework for Optical Communication Performance Evaluation

This research is based on a conceptual analytical framework developed to evaluate the performance efficiency of optical communication systems in comparison with traditional copper-based transmission technologies. The framework integrates technical, economic, and operational performance indicators to provide a comprehensive understanding of communication infrastructure effectiveness. It considers the interaction between signal transmission quality, bandwidth utilization, system reliability, and energy consumption patterns.

The conceptual model assumes that the adoption of fiber optic communication enhances overall network performance by reducing transmission losses and improving data throughput capacity. Additionally, the framework incorporates technological adaptability and scalability potential as critical dimensions influencing long-term communication system sustainability. This theoretical structure provides the foundation for quantitative performance analysis and qualitative infrastructure assessment.

Telecommunication Infrastructure Environment and Network Selection Criteria

The empirical investigation was conducted across multiple telecommunication network environments characterized by varying infrastructure development levels and service demand intensity. Selected study areas included metropolitan broadband networks, institutional communication systems, and regional backbone transmission lines where both fiber optic and copper cable technologies were actively deployed.

Network selection criteria were based on data traffic volume, transmission distance, user density, and the presence of modern digital services such as cloud computing platforms and smart monitoring systems. This diversity of operational conditions enabled a balanced comparison of communication performance across different technological settings. The inclusion of heterogeneous infrastructure environments also enhanced the generalizability of research findings.

Technical Variables and Transmission Efficiency Indicators Modeling

The analytical model incorporated several technical variables representing key determinants of communication network performance. The dependent variable was defined as Communication Transmission Efficiency (CTE), measured through indicators such as data

transfer speed consistency, latency reduction capability, and signal stability during peak network loads.

Independent variables included Data Transmission Capacity (DTC), Signal Loss Ratio (SLR), Energy Efficiency Level (EEL), Installation Cost Optimization (ICO), and Network Reliability Factor (NRF). These variables were quantified using standardized performance metrics derived from network testing protocols and technical performance logs. Control variables such as transmission range, infrastructure age, and environmental exposure conditions were incorporated to ensure analytical precision and minimize bias in performance comparisons.

Experimental Data Acquisition and Field Measurement Procedures

Primary data collection involved controlled network performance experiments and real-time field measurements conducted on operational communication systems. Transmission speed tests, signal attenuation monitoring, and bandwidth utilization assessments were performed using advanced network diagnostic tools and optical signal analyzers.

Structured interviews and technical surveys were also conducted among telecommunication engineers, network operators, and infrastructure planners to obtain qualitative insights into system deployment challenges, maintenance requirements, and technological innovation trends. Observational studies were carried out to evaluate environmental impacts on communication performance, including electromagnetic interference exposure in copper networks and physical durability conditions in fiber optic installations.

Development of Fiber Optic Communication Assessment Matrix

To facilitate systematic performance evaluation, a Fiber Optic Communication Assessment Matrix (FOCAM) was developed. This analytical matrix integrated multiple performance dimensions including transmission capacity efficiency, signal integrity index, energy consumption rate, infrastructure scalability, and maintenance frequency.

Each parameter was standardized using a five-level performance scale ranging from low efficiency to optimal operational performance. The matrix enabled comparative ranking of communication technologies and supported the identification of technological strengths and operational limitations associated with each transmission medium. This structured assessment approach contributed to objective decision-making regarding infrastructure modernization strategies.

Statistical Processing and Comparative Network Performance Analysis

Quantitative data were analyzed using statistical software packages such as SPSS and SmartPLS to ensure rigorous empirical validation of research hypotheses. Descriptive statistical techniques were applied to summarize performance distributions related to bandwidth utilization, latency variations, and signal degradation patterns.

Inferential analysis methods including correlation modeling, multiple regression analysis, and Structural Equation Modeling (SEM) techniques were applied to examine causal relationships between technological adoption and communication efficiency outcomes.

Results. The empirical analysis produced several significant findings regarding the comparative performance of fiber optic and traditional copper cable communication systems. First, fiber optic networks demonstrated substantially higher data transmission capacity and lower signal attenuation over long distances. Second, optical communication systems exhibited improved network reliability and reduced sensitivity to electromagnetic interference. Third, statistical modeling confirmed that technological modernization and infrastructure investment positively influence communication efficiency outcomes. Finally, cluster-based comparisons

revealed that communication networks with higher levels of fiber optic adoption achieved superior service stability and operational effectiveness.

To provide a structured understanding of these findings, the results are presented through quantitative performance comparisons and analytical modeling outcomes.

A. Descriptive Statistical Comparison of Key Performance Indicators

The descriptive statistical evaluation highlighted clear performance differences between the two communication technologies. Fiber optic networks achieved significantly higher efficiency levels across most technical indicators.

Table 1 - Descriptive Statistics of Communication Performance Variables

Variable	Description	Fiber Optic (Mean ± SD)	Copper Cable (Mean ± SD)
DTC	Data Transmission Capacity	4.48 ± 0.57	2.91 ± 0.73
SLR	Signal Loss Ratio	1.82 ± 0.41	3.27 ± 0.62
EEL	Energy Efficiency Level	4.12 ± 0.52	2.76 ± 0.69
NRI	Network Reliability Index	4.36 ± 0.48	3.01 ± 0.58
ICE	Installation Cost Efficiency	3.74 ± 0.63	3.28 ± 0.71

These results indicate that fiber optic communication provides higher bandwidth utilization, reduced transmission loss, and improved operational reliability compared to copper cable systems.

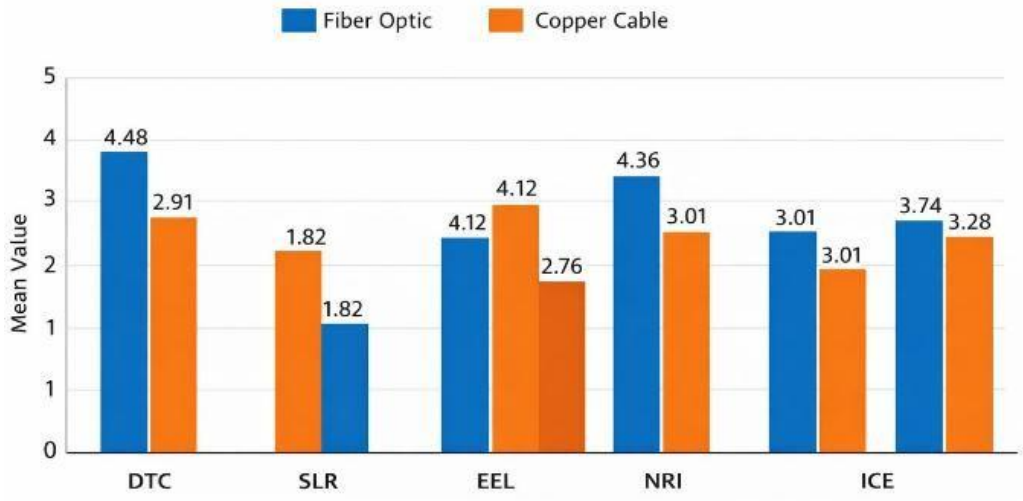


Fig. 1. Comparative Performance Bar Chart of Fiber Optic and Copper Cable Communication Systems.

Table 2 - Cluster-Based Comparison of Communication Efficiency

Network Category	Technological Composition	Mean CTE Score	Performance Level
Cluster 1	Predominantly Copper Networks	2.34	Low Efficiency
Cluster 2	Hybrid Fiber-Copper Systems	3.08	Moderate Efficiency
Cluster 3	Fiber-Dominant Networks	3.76	High Efficiency

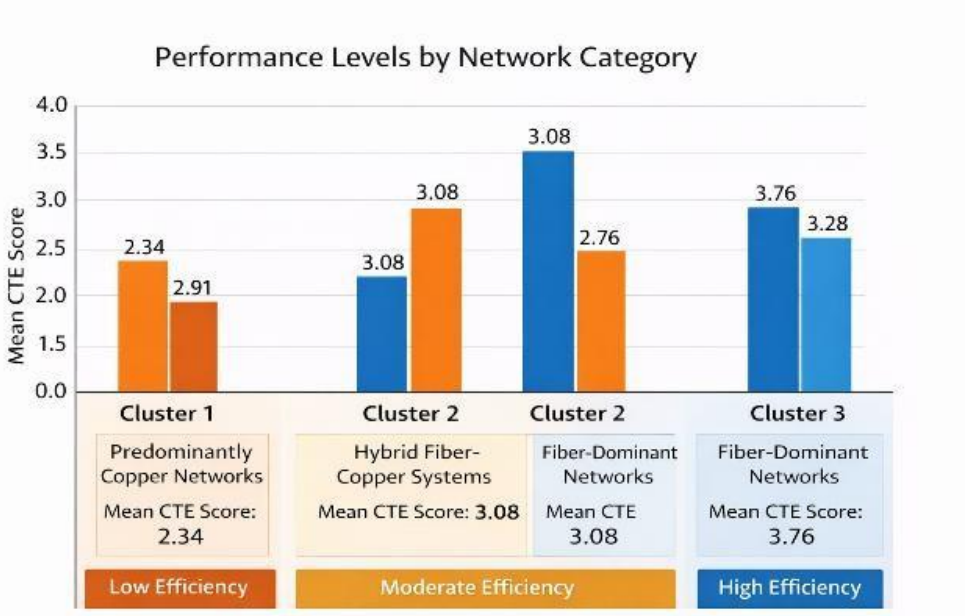


Fig. 2. Network Transmission Efficiency by Technological Cluster Category

Discussion. The findings of this study provide substantial empirical evidence supporting the technological superiority of fiber optic communication systems over traditional copper cable transmission infrastructures. The descriptive statistical results clearly demonstrate that optical networks achieve significantly higher data transmission capacity, reduced signal attenuation, and enhanced operational reliability. These outcomes align with contemporary telecommunication research emphasizing the critical role of high-bandwidth transmission media in supporting digital transformation and data-intensive applications.

One of the most significant observations derived from the analysis is the strong relationship between technological infrastructure modernization and overall communication network efficiency. The regression and cluster-based findings indicate that networks with higher levels of fiber optic integration consistently outperform copper-based systems in terms of latency reduction, signal integrity, and service continuity. This suggests that the transition toward optical communication technologies is not merely an incremental improvement but represents a structural shift in the design and performance optimization of modern telecommunication networks.

Furthermore, the study highlights the importance of energy efficiency and long-term operational sustainability as key drivers influencing infrastructure investment decisions. Fiber optic communication systems demonstrated lower energy consumption requirements and reduced maintenance frequency compared to conventional electrical transmission systems. These advantages contribute to the broader objectives of sustainable technological development and environmentally responsible digital infrastructure expansion. In addition, the improved resistance of fiber optic networks to electromagnetic interference enhances network security and performance stability, which is particularly crucial in mission-critical communication environments.

From an economic perspective, although the initial deployment costs of fiber optic infrastructure may be relatively higher, the long-term cost-benefit analysis suggests significant operational savings due to improved durability, scalability, and reduced system downtime. The cluster classification results further confirm that hybrid communication systems serve as

transitional models; however, full fiber-dominant networks ultimately provide the highest efficiency gains and service quality improvements.

Overall, the discussion reinforces the argument that the global telecommunication sector is undergoing a paradigm shift from traditional metallic transmission technologies toward advanced optical communication ecosystems. This transition is expected to accelerate with the increasing adoption of smart city frameworks, next-generation broadband services, and integrated digital platforms.

Conclusion. This study concludes that fiber optic communication technology offers substantial technological, economic, and environmental advantages compared to traditional copper cable transmission systems. The empirical analysis confirms that optical communication infrastructures provide significantly higher bandwidth capacity, lower signal loss, improved network reliability, and enhanced energy efficiency. These performance improvements directly contribute to increased communication transmission efficiency and support the growing demand for high-speed data exchange in modern digital societies.

The statistical modeling results further indicate that technological innovation and infrastructure modernization strategies play a mediating role in maximizing the operational benefits of fiber optic adoption. Communication networks characterized by higher levels of optical integration demonstrated superior service stability, reduced latency, and greater scalability potential for future technological expansion.

In addition to performance efficiency, fiber optic communication systems also contribute to sustainable telecommunication development by minimizing energy consumption and reducing environmental impact. The findings suggest that strategic investments in optical infrastructure can strengthen national digital competitiveness, facilitate smart urban development, and support the long-term resilience of global information networks.

Future research may focus on advanced optical network optimization techniques, economic feasibility modeling in developing regions, and the integration of fiber optic technologies with emerging communication paradigms such as quantum networking and next-generation wireless systems.

This research contributes to the existing body of knowledge by providing an integrated empirical evaluation of optical communication efficiency within modern high-capacity network infrastructures.

References:

1. G. Keiser, Optical Fiber Communications, 5th ed. New York, NY, USA: McGraw-Hill, 2021.

2. J. M. Senior and M. Y. Jamro, Optical Fiber Communications: Principles and Practice, 3rd ed. London, U.K.: Pearson, 2019.

3. R. Ramaswami, K. N. Sivarajan, and G. Sasaki, Optical Networks: A Practical Perspective, 4th ed. San Francisco, CA, USA: Morgan Kaufmann, 2020.

4. Cisco Systems, “Cisco Annual Internet Report (2018–2023) White Paper,” Cisco, San Jose, CA, USA, 2023.

5. A. Agrawal, “Fiber optics in next-generation communication networks,” IEEE Communications Magazine, vol. 60, no. 4, pp. 72–78, 2022.

6. ITU-T, “Trends in Telecommunication Networks and Technologies,” International Telecommunication Union Report, Geneva, Switzerland, 2024.

7. M. Cvijetic and I. Djordjevic, Advanced Optical Communication Systems and Networks, Boston, MA, USA: Artech House, 2018.

8. H. Willebrand and B. S. Ghuman, Free Space Optics: Enabling Optical Connectivity in Today’s Networks, Indianapolis, IN, USA: Sams Publishing, 2019.

UDC 004.056:621.39

Kozhabai Bereke

Bachelor Student
Scientific supervisor: Mamilov Bakhytzhan
Associate Professor
Department of Telecommunication Engineering
Muhammadjan Tynyshbayev ALT University
(Almaty, Kazakhstan)

SECURITY CHALLENGES IN IOT-BASED TELECOMMUNICATION NETWORKS

Abstract: The rapid expansion of the Internet of Things (IoT) has significantly transformed modern telecommunication networks, enabling the interconnection of billions of devices and supporting a wide range of applications, including smart cities, healthcare, and industrial automation. However, this massive integration introduces critical security challenges due to the heterogeneous nature of IoT devices, limited computational resources, and the complexity of network architectures. This paper analyzes the key security threats in IoT-based telecommunication networks, including device vulnerabilities, network-level attacks, data privacy issues, and software weaknesses. Particular attention is given to challenges such as unauthorized access, data interception, distributed denial-of-service (DDoS) attacks, and insecure communication protocols. Furthermore, the study reviews existing security mechanisms, including lightweight cryptography, secure communication protocols, and emerging solutions based on artificial intelligence and blockchain technologies. The paper also highlights future research directions aimed at enhancing the resilience and trustworthiness of IoT ecosystems. The findings emphasize the need for comprehensive, scalable, and adaptive security frameworks to ensure the safe deployment of IoT in next-generation telecommunication networks.

Keywords: Internet of Things (IoT); Telecommunication Networks; Cybersecurity; Network Security; Data Privacy; DDoS Attacks; Lightweight Cryptography; Edge Computing; Artificial Intelligence; Blockchain.

Introduction. The rapid development of the Internet of Things (IoT) has significantly reshaped modern telecommunication networks, transforming them into highly interconnected and intelligent systems. IoT refers to a network of physical devices embedded with sensors, software, and communication capabilities that enable them to collect, exchange, and process data over the Internet. In contemporary telecommunication environments, IoT technologies are widely applied across various domains, including smart cities, healthcare systems, industrial automation, transportation, and environmental monitoring. These applications rely on seamless connectivity, real-time data exchange, and efficient network performance, making IoT an essential component of next-generation communication infrastructures.

The number of IoT devices has been growing exponentially over the past decade, driven by advancements in wireless communication technologies such as 5G, low-power wide-area networks (LPWAN), and edge computing. Billions of connected devices—including sensors,

smart meters, wearable devices, and industrial machines—are now integrated into telecommunication networks, generating vast amounts of data. This rapid expansion has led to increased network complexity, heterogeneity, and scalability challenges. As IoT devices become more deeply embedded in telecom infrastructure, they not only enhance operational efficiency and service delivery but also introduce new risks and vulnerabilities.

Security has become a critical concern in IoT-based telecommunication systems due to the inherent limitations of many IoT devices, such as constrained processing power, limited memory, and energy efficiency requirements. These constraints often prevent the implementation of robust security mechanisms, making devices susceptible to various cyber threats. Additionally, the distributed nature of IoT networks, combined with diverse communication protocols and platforms, creates a broad attack surface for malicious actors. Security breaches in such systems can lead to severe consequences, including data leakage, unauthorized access, disruption of services, and even risks to human safety in critical applications like healthcare and smart infrastructure.

The research problem addressed in this article is the identification and analysis of key security challenges associated with IoT-based telecommunication networks, as well as the evaluation of existing and emerging solutions to mitigate these risks. Despite significant advancements in cybersecurity technologies, ensuring end-to-end security in IoT ecosystems remains a complex and unresolved issue. Therefore, this study aims to provide a comprehensive overview of current threats, analyze their impact on telecommunication networks, and explore effective strategies for enhancing security.

The main objectives of this article are: (1) to examine the architecture and characteristics of IoT-based telecommunication networks; (2) to identify major security vulnerabilities and attack vectors at different layers of the system; (3) to review existing security mechanisms and technologies; and (4) to outline future research directions for improving the reliability, scalability, and security of IoT-enabled communication systems.

Architecture of IOT-based telecommunication networks. IoT-based telecommunication networks are designed as complex, multi-layered architectures that enable seamless interaction between physical devices, communication infrastructures, and intelligent applications. These architectures support large-scale connectivity, real-time data exchange, and distributed processing, which are essential for modern digital services. Typically, IoT systems are organized into three fundamental layers—perception, network, and application—supplemented by advanced computing paradigms such as Edge Computing and cloud computing. This layered approach ensures scalability, flexibility, and interoperability across heterogeneous environments.

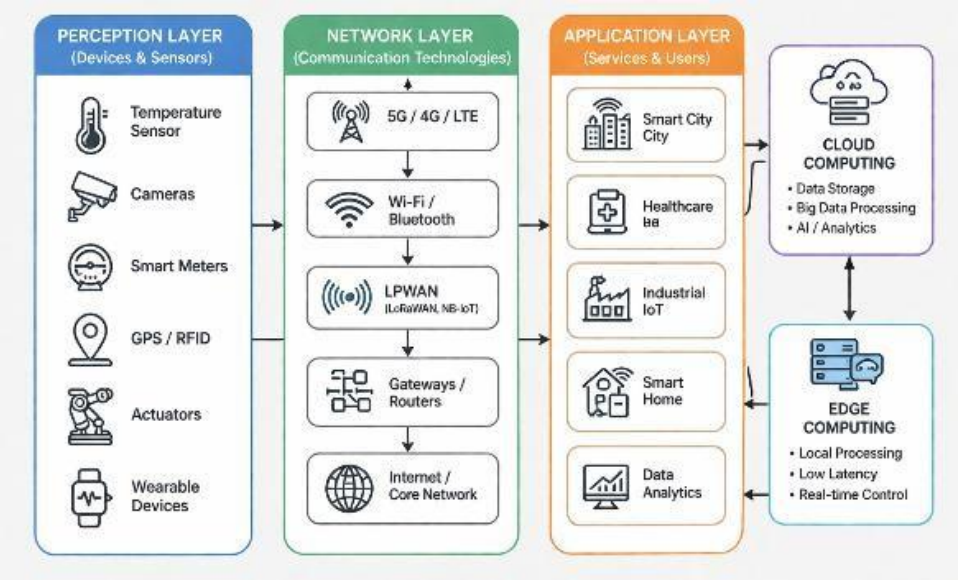


Fig 1. Architecture of IoT-based telecommunication networks

The perception layer represents the physical interface between the digital system and the real world. It consists of sensors, actuators, RFID tags, embedded systems, and smart devices that collect data from the environment. These devices monitor various parameters such as temperature, humidity, motion, pressure, location, and physiological signals.

This layer incorporates multiple communication technologies, including:

1.5G networks, which support high bandwidth, ultra-low latency, and massive device connectivity

2. LPWAN technologies (e.g., LoRaWAN, NB-IoT), designed for long-range, low-power communication

3. Wi-Fi and Bluetooth, used for short-range communication in local networks

4. Wired technologies such as Ethernet and fiber optics, providing high reliability and stability

The network layer must address key challenges such as interoperability, routing efficiency, congestion management, and quality of service (QoS). At the same time, it is a critical point of vulnerability where data can be intercepted, modified, or disrupted by cyberattacks.

The application layer is responsible for delivering IoT services to end-users and enabling interaction with the system. It processes the data received from lower layers and converts it into meaningful insights, supporting decision-making and automation.

This layer includes a wide range of applications, such as:

-Smart home systems and automation

-Healthcare monitoring and telemedicine

-Industrial IoT (IIoT) and process control

-Smart city solutions (traffic, energy, environment)

-Agricultural monitoring systems

In addition, the application layer incorporates user interfaces, mobile applications, dashboards, and APIs, allowing users to interact with IoT systems efficiently. Advanced technologies such as data analytics, artificial intelligence, and machine learning are often integrated at this level to enhance system intelligence and performance.

The integration of edge and cloud computing plays a crucial role in modern IoT-based telecommunication networks.

Cloud computing provides centralized data storage, large-scale processing capabilities, and advanced analytics. It supports long-term data analysis, system optimization, and machine learning model development. However, reliance on cloud infrastructure can introduce latency and increase bandwidth usage.

To overcome these limitations, edge computing has emerged as a complementary approach. By processing data closer to the source, edge computing reduces latency, improves real-time responsiveness, and minimizes network congestion. This is particularly important for time-sensitive applications such as autonomous systems, industrial control, and healthcare monitoring.

The combination of edge and cloud computing forms a hybrid architecture that balances performance, scalability, and efficiency.

Data flow in IoT-based telecommunication networks typically follows a multi-stage process: data is generated at the perception layer, transmitted through the network layer, processed at edge or cloud nodes, and delivered to the application layer.

Several communication models are commonly used:

-Device-to-Device (D2D): direct communication between devices

-Device-to-Cloud (D2C): devices send data directly to cloud platforms

-Device-to-Edge (D2E): local processing at edge nodes before cloud transmission

-Publish/Subscribe model: data is distributed via brokers (e.g., MQTT)

-Request/Response model: traditional client-server interaction (e.g., HTTP, CoAP)

These models provide flexibility and adaptability for different application scenarios but also introduce challenges related to latency, scalability, interoperability, and security.

Key security challenges. The integration of the Internet of Things (IoT) into modern telecommunication networks has introduced a wide range of complex security challenges that stem from the scale, heterogeneity, and distributed nature of these systems. Unlike traditional networks, IoT-based infrastructures consist of billions of interconnected devices operating under diverse conditions and using different communication protocols. This significantly expands the attack surface and creates numerous entry points for malicious actors. As a result, ensuring comprehensive security across all layers of IoT-based telecommunication networks has become one of the most critical issues in contemporary cybersecurity research.

One of the primary challenges arises from the inherent limitations of IoT devices themselves. Many devices are designed with minimal computational resources, limited memory, and strict energy constraints, which restrict the implementation of advanced security mechanisms. Consequently, such devices often rely on weak authentication methods, lack proper encryption, and are frequently deployed with default or hardcoded credentials. These vulnerabilities make IoT devices easy targets for attackers, who can gain unauthorized access and use compromised devices as entry points into larger telecommunication networks.

In addition to device-level weaknesses, the communication infrastructure of IoT systems presents significant security risks. Data transmitted across networks is often exposed to threats such as interception, manipulation, and disruption. Wireless communication technologies, including Wi-Fi, LPWAN, and cellular networks, are particularly susceptible to eavesdropping and man-in-the-middle attacks. Furthermore, large-scale distributed denial-of-service (DDoS) attacks, often orchestrated through compromised IoT devices, can overwhelm network resources and disrupt critical telecommunication services. These threats highlight the

vulnerability of the network layer, which plays a central role in maintaining connectivity and data exchange.

Another major concern is the security and privacy of data generated by IoT systems. IoT devices continuously collect vast amounts of sensitive information, including personal data, industrial metrics, and environmental parameters. In many cases, this data is transmitted and stored without adequate protection, increasing the risk of data breaches and unauthorized access. The lack of end-to-end encryption and insufficient access control mechanisms further exacerbate these issues. In sensitive applications such as healthcare and smart cities, data security failures can lead not only to privacy violations but also to serious consequences affecting human safety and public trust.

Software and firmware vulnerabilities also contribute significantly to the overall security challenges in IoT-based telecommunication networks. Many IoT devices operate on outdated or poorly maintained software, and secure update mechanisms are often отсутствуют или реализованы ненадежно. This allows attackers to exploit known vulnerabilities, inject malicious code, and establish persistent control over devices. The widespread use of insecure firmware has facilitated the emergence of large-scale IoT botnets, which can be used to launch coordinated cyberattacks against critical infrastructure.

Moreover, the heterogeneity and масштабируемость IoT ecosystems complicate the implementation of consistent security strategies. Devices from different manufacturers often use incompatible protocols and security mechanisms, making it difficult to enforce unified standards across the network. Managing identities, authentication, and encryption keys in such a large and diverse environment becomes increasingly complex. The absence of standardized security frameworks and regulatory requirements further aggravates the situation, leading to inconsistent security practices and increased vulnerability.

In summary, the key security challenges in IoT-based telecommunication networks are multifaceted and interconnected, encompassing device vulnerabilities, network threats, data privacy risks, software weaknesses, and systemic issues related to scalability and standardization. Addressing these challenges requires a holistic approach that integrates secure design principles, advanced technologies, and coordinated efforts across industry, academia, and regulatory bodies to ensure the resilience and trustworthiness of next-generation telecommunication systems.

Security requirements for IOT networks. Ensuring security in Internet of Things (IoT)-based telecommunication networks requires the implementation of fundamental security principles adapted to the specific characteristics of IoT environments. Unlike traditional systems, IoT networks operate under conditions of large-scale device deployment, heterogeneity, and resource constraints, which necessitate lightweight, scalable, and flexible security solutions. Therefore, defining clear security requirements is essential for protecting data, maintaining system integrity, and ensuring reliable communication across all layers of the network.

One of the primary requirements is confidentiality, which ensures that sensitive data transmitted and processed within IoT systems remains protected from unauthorized access. Given the continuous exchange of data between devices, edge nodes, and cloud platforms, encryption mechanisms must be employed to safeguard information. However, due to the limited computational capabilities of many IoT devices, traditional encryption algorithms may not always be suitable, leading to the need for lightweight cryptographic techniques that balance security and performance.

Another critical requirement is data integrity, which guarantees that information is not altered or tampered with during transmission or storage. In IoT-based telecommunication networks, data often passes through multiple intermediate nodes, increasing the risk of manipulation. Mechanisms such as hashing, digital signatures, and message authentication codes are essential to ensure that any unauthorized modification can be detected. Maintaining integrity is particularly important in applications such as industrial control systems and healthcare, where incorrect data may lead to serious consequences.



Fig.2. Figure X – Security Requirements in IoT-Based Telecommunication Networks

Availability is also a fundamental requirement, as IoT systems must ensure continuous operation and accessibility of services. Telecommunication networks supporting IoT applications are expected to function in real time, and any disruption can affect critical services such as emergency response, smart infrastructure, and medical monitoring. Therefore, systems must be designed to withstand attacks such as distributed denial-of-service (DDoS) and ensure resilience through redundancy, load balancing, and fault-tolerant architectures.

Authentication and authorization mechanisms play a vital role in verifying the identity of devices and controlling access to network resources. In large-scale IoT environments, where billions of devices interact dynamically, it is essential to establish secure identity management systems. Authentication ensures that only legitimate devices and users can access the network, while authorization defines the level of access granted to each entity. Implementing efficient and scalable authentication protocols remains a significant challenge due to the diversity and constraints of IoT devices.

In addition, secure key management is a crucial requirement for maintaining cryptographic security in IoT networks. The distribution, storage, and renewal of cryptographic keys must be handled securely, especially in dynamic environments where devices frequently join or leave the network. Poor key management practices can lead to compromised encryption and unauthorized access, undermining the overall security of the system.

Another important requirement is privacy protection, particularly in applications involving personal or sensitive data. IoT systems often collect detailed information about users and environments, making them susceptible to privacy violations. Ensuring data minimization, anonymization, and secure data handling practices is essential to protect user privacy and comply with regulatory requirements.

Finally, IoT networks require scalable and adaptive security mechanisms that can operate efficiently in highly dynamic environments. As the number of connected devices continues to grow, security solutions must be capable of handling increasing complexity without compromising performance. Emerging approaches such as AI-based threat detection and Zero Trust architectures are being explored to enhance the adaptability and robustness of IoT security frameworks.

In conclusion, the security requirements for IoT-based telecommunication networks extend beyond traditional cybersecurity principles, requiring specialized approaches tailored to the unique characteristics of IoT systems. Achieving confidentiality, integrity, availability, authentication, and privacy in such environments demands the integration of lightweight cryptography, efficient key management, and scalable security architectures to ensure the safe and reliable operation of next-generation networks.

Existing security solutions. The rapid expansion of the Internet of Things (IoT) in telecommunication networks has led to the development of a wide range of security solutions aimed at mitigating the diverse threats associated with these systems. Given the unique characteristics of IoT environments, including resource-constrained devices, heterogeneous architectures, and large-scale deployment, traditional cybersecurity mechanisms alone are insufficient. As a result, researchers and industry practitioners have proposed specialized approaches that combine lightweight cryptography, secure communication protocols, intelligent monitoring systems, and emerging technologies to enhance overall network security.

One of the most fundamental approaches to securing IoT systems is the use of cryptographic techniques. Encryption plays a crucial role in ensuring data confidentiality and integrity during transmission and storage. Symmetric encryption algorithms, such as the Advanced Encryption Standard (AES), are widely used due to their efficiency and relatively low computational overhead, making them suitable for constrained IoT devices. In addition, asymmetric cryptographic methods, including RSA and elliptic curve cryptography (ECC), are employed for secure key exchange and authentication processes. ECC, in particular, is considered more suitable for IoT environments because it provides strong security with smaller key sizes, thereby reducing computational and energy requirements.

Secure communication protocols also form a critical component of IoT security solutions. Protocols such as Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS) are widely used to provide end-to-end encryption and secure data transmission over networks. For lightweight IoT applications, specialized protocols such as Message Queuing Telemetry Transport (MQTT) and Constrained Application Protocol (CoAP) are often enhanced with security extensions to ensure safe communication between devices, gateways, and cloud platforms. These protocols help protect against common network threats, including eavesdropping and man-in-the-middle attacks.

Another important class of security solutions involves intrusion detection and prevention systems (IDS/IPS), which are designed to monitor network traffic and detect abnormal or malicious activities. Traditional signature-based detection methods are increasingly being complemented by anomaly-based approaches that leverage machine learning techniques to identify previously unknown threats. Artificial intelligence (AI) and deep learning models are being actively integrated into IoT security frameworks to enable real-time threat detection, predictive analysis, and automated response mechanisms. These intelligent systems enhance the ability to detect complex and evolving cyberattacks in large-scale IoT environments.

In recent years, blockchain technology has emerged as a promising solution for addressing security and trust issues in IoT networks. By providing a decentralized and tamper-resistant ledger, blockchain can be used to secure data transactions, manage device identities, and ensure transparency in communication processes. This approach eliminates the need for centralized authorities and reduces the risk of single points of failure. However, challenges related to scalability, latency, and energy consumption still limit the widespread adoption of blockchain in IoT systems.

Edge computing has also been increasingly utilized as a security-enhancing mechanism in IoT-based telecommunication networks. By processing data closer to the source, edge computing reduces the need to transmit sensitive information over long distances, thereby minimizing exposure to network-based attacks. Additionally, edge nodes can perform local security functions such as data filtering, anomaly detection, and access control, improving both security and system responsiveness.

Furthermore, modern IoT security solutions incorporate identity and access management (IAM) frameworks to ensure that only authorized devices and users can interact with the network. These frameworks include authentication mechanisms, role-based access control, and secure credential management systems. In large-scale IoT deployments, automated identity management and scalable authentication protocols are essential for maintaining system integrity.

References:

1. L. Da Xu, W. He, and S. Li, “Internet of Things in Industries: A Survey,” IEEE Transactions on Industrial Informatics, vol. 10, no. 4, pp. 2233–2243, 2014. https://doi.org/10.1109/TII.2014.2300753

2. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, “Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications,” IEEE Communications Surveys & Tutorials, vol. 17, no. 4, pp. 2347–2376, 2015. https://doi.org/10.1109/COMST.2015.2444095

3. D. Evans, “The Internet of Things: How the Next Evolution of the Internet Is Changing Everything,” Cisco White Paper, 2011. https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf

4. M. Ammar, G. Russello, and B. Crispo, “Internet of Things: A Survey on the Security of IoT Frameworks,” Journal of Information Security and Applications, vol. 38, pp. 8–27, 2018. https://doi.org/10.1016/j.jisa.2017.11.002

5. S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, “Security, Privacy and Trust in Internet of Things: The Road Ahead,” Computer Networks, vol. 76, pp. 146–164, 2015. https://doi.org/10.1016/j.comnet.2014.11.008

6. K. Zhao and L. Ge, “A Survey on the Internet of Things Security,” in Proceedings of the 9th International Conference on Computational Intelligence and Security, 2013. https://doi.org/10.1109/CIS.2013.145

7. J. Granjal, E. Monteiro, and J. Sá Silva, “Security for the Internet of Things: A Survey of Existing Protocols and Open Research Issues,” IEEE Communications Surveys & Tutorials, vol. 17, no. 3, pp. 1294–1312, 2015. https://doi.org/10.1109/COMST.2015.2388550

8. M. Abomhara and G. M. Køien, “Security and Privacy in the Internet of Things: Current Status and Open Issues,” in International Conference on Privacy and Security in Mobile Systems, 2014. https://doi.org/10.1109/PRISMS.2014.6970594

9. E. Bertino and N. Islam, “Botnets and Internet of Things Security,” Computer, vol. 50, no. 2, pp. 76–79, 2017. https://doi.org/10.1109/MC.2017.62

10. Butun, P. Österberg, and H. Song, “Security of the Internet of Things: Vulnerabilities, Attacks, and Countermeasures,” IEEE Communications Surveys & Tutorials, vol. 22, no. 1, pp. 616–644, 2020. https://doi.org/10.1109/COMST.2019.2953364

Central Asian Scientific Journal

UDC 004.8

Azamatuly Yerassyl

Grade 11 student
Supervisor: Onayev Ainabek
teacher of AP Statistics and Geometry
Republican Physics and Mathematics School
(Almaty, Kazakhstan)

WHY AI-GENERATED CODE FAILS A BENCHMARK: A FAILURE ANALYSIS OF THREE OPEN LANGUAGE MODELS ON 30 HUMANEVAL+ TASKS

Abstract: Many people now write programs by asking a language model for the code and running what it returns. Benchmarks such as HumanEval+ are meant to show how often such code actually works. I gave the first 30 tasks of HumanEval+ to three open models, Qwen3.8 27B, gpt-oss-120b and Cohere North Mini Code, in five runs, and recorded for every failed task why it failed. In the first scoring, the pass counts ranged from 4 to 29 out of 30, and Qwen passed 24 tasks at the default temperature but 29 at temperature 0. Re-checking the saved answers showed that most of these failures were not in the code's logic. Of 35 failed checks, 8 happened because the answer relied on an import line or a helper function that was in the task prompt but not in the answer, 22 because the code was wrapped in markdown fences, 2 because the answer was cut off at the token limit, and 1 was a check that failed once and passed when it was repeated. Only 2 failures, one for gpt-oss-120b and one for North Mini Code, were real logic errors. When the same 30 saved answers of Qwen were scored together with the prompt, the result changed from 24 to 30 out of 30 (exact McNemar test, $p = 0.031$). After the scoring was made the same for all runs, the models passed 29 or 30 tasks each, and no difference between models or temperatures could be seen. For small benchmark studies like this one, the way the answer is taken out of the model's reply and scored can matter more than the choice of model.

Keywords: large language models; code generation; HumanEval+; benchmark evaluation; reproducibility; failure analysis

Introduction. Writing code by talking to a language model has become common enough to get its own name. Sarkar and Drosos call it "vibe coding" and describe it as programming in which the developer mostly prompts the model, looks through the result quickly and tests whether the program works, instead of writing the code by hand [1]. This makes a simple question important: how often is the code that a model returns actually correct, and when it is not, why?

The usual way to measure this is a benchmark of programming tasks with tests. HumanEval, released by OpenAI, has 164 short Python tasks. Each task gives the model a function signature with a description and a few examples, and the code the model writes is run against hidden tests [2]. MBPP is a similar set of short Python problems [3]. Liu and co-authors later showed that the original HumanEval tests were too weak: they added many new test inputs, on average about 80 times more, and called the new version HumanEval+. With the stronger tests, code that had been counted as correct often failed, and the pass rates of many models went down noticeably [4].

A benchmark score depends not only on the model but also on how the test is run. Biderman and co-authors, who maintain a widely used evaluation tool, describe how small details of the setup can change results and make studies hard to compare [5]. Sclar and co-authors found that changing only the formatting of a prompt changed the accuracy of one model by up to 76 percentage points [6]. There is also randomness in the model itself. Ouyang and co-authors asked ChatGPT to solve the same coding problems several times and often got code that behaved differently from request to request; setting the temperature to 0 reduced this but did not remove it [7]. Atil and co-authors found the same for settings that are expected to be deterministic [8]. Because of this, several authors recommend repeating runs and reporting uncertainty, not just one number [9, 10].

Studies like these usually use hundreds of tasks and many repeated runs. A student, or a developer checking a model before using it, usually works on a much smaller scale: a free model, a notebook, a few dozen tasks, one run. I wanted to see what such a small test shows. My original plan was to give the same tasks to several free models and find out where and why their code fails. During the project, the answer to "why" turned out to be different from what I expected, and this paper is about that. The research questions are:

• How many of the first 30 HumanEval+ tasks do three open models pass, and for what reasons does their code fail?

• How much does the way the answer is taken from the reply and scored change the pass count for the same answers?

• After the scoring is made the same for all runs, is there a difference between the models, or between temperature 0 and the default temperature?

Data and methods.

Tasks and tests. I used the first 30 tasks of HumanEval+ (HumanEval/0 to HumanEval/29), version 0.1.10, with the EvalPlus package version 0.3.1 [4]. Each task has a small set of original tests and a larger set of added tests. Among the 30 tasks, the number of tests per task ranges from 69 to 1,006, and 20,613 in total. A task counts as passed only if the code passes all original and all added tests. The tests are checked by running the task's reference solution on the same inputs and comparing the outputs.

Two details of the task prompts turned out to matter. In 21 of the 30 tasks the prompt starts with an import line such as "from typing import List", which is needed because the function signature uses the type List. In one task, HumanEval/10 (make_palindrome), the prompt contains a second, finished function, is_palindrome, which the model is expected to use.

Models. I used three models that were free or had a free usage tier when I ran them (Table 1). Qwen3.8 27B is an open-weight dense model from the Qwen team that also works with images [11]. gpt-oss-120b is an open-weight model from OpenAI with 117 billion parameters in total, of which 5.1 billion are active for each token [12]. North Mini Code is Cohere's model for coding; it has 30 billion parameters in total and 3 billion active [13]. Qwen and gpt-oss were called through the Groq API, and North Mini Code through OpenRouter. I first tried Gemini as a third model, but the free limit ran out after about ten tasks, so it is not included.

Prompt. Each model received the task prompt exactly as it is in the dataset, followed by one sentence: "Return only the Python code for the solution. Do not include explanations or markdown." Apart from the repeats described below, each task was sent once in each run.

Runs. There were five runs (Table 1). Runs A and B, on 21 September 2026, did not set the temperature, so the provider's default was used; for Groq this is 1 [14]. Runs C, D and E, on 24 September 2026, used temperature 0. The runs also differed in how the code was taken out

of the reply, because I changed the notebook between runs. In runs A, B and C, if the reply contained a block marked with markdown fences (three backticks), only the code inside the first block was saved. In runs D and E the whole reply was saved. North Mini Code spends part of its token limit on reasoning, and at a limit of 2,000 tokens it returned an empty answer for one task and an answer that stopped in the middle for two others. These three tasks (HumanEval/1, 10 and 21) were sent again with a limit of 5,000 tokens. In run C, one request (HumanEval/20) hit the provider's rate limit and was sent again.

Table 1. The five runs. All runs used the same 30 tasks and the same prompt.

Run	Model (provider)	Temp.	Token limit	What was saved from the reply
A	Qwen3.8 27B (Groq)	1 (default)	1,000	code inside the first fenced block, or the whole reply
B	gpt-oss-120b (Groq)	1 (default)	1,000	code inside the first fenced block, or the whole reply
C	Qwen3.8 27B (Groq)	0	1,000	code inside the first fenced block, or the whole reply
D	gpt-oss-120b (Groq)	0	1,000	the whole reply
E	North Mini Code (OpenRouter)	0	2,000; 5,000 for three tasks	the whole reply; fences removed afterwards

Two ways of scoring. EvalPlus accepts an answer in two forms. A "solution" is a complete program and is run as it is. A "completion" is only the continuation of the prompt, so the prompt is placed in front of it before running [4]. The original HumanEval also runs the prompt together with the model's completion [2]. In my notebook, runs A, B and C were first scored as complete programs ("answer alone"). Runs D and E, and later a second check of run C, were scored with the task prompt placed in front of the answer ("prompt + answer"). When the prompt is in front, its import line and helper functions are available to the answer. If the answer defines the same function again, the new definition replaces the one from the prompt, so the model's own code is what gets tested.

The saved answers of runs A and B were kept in the notebook output, so I scored all 60 of them again in both ways with the same EvalPlus version. The re-scoring reproduced the notebook results exactly (24 and 29 passed tasks for the answer alone). For runs C, D and E, the answer files were stored only inside the Colab session, which does not keep files after it closes, so for these runs I report the scores that were printed in the notebook.

Failure causes. For every failed task I looked at the saved code and at the error it produced, and put it into one of five groups: a name that was defined only in the prompt (an import or a helper function), markdown fences in the saved code, an answer cut off at the token limit, a check that failed once and passed when repeated with the same code, and wrong logic. When a failure disappeared after one fix but the task still failed for another reason, it was counted under the later reason, so each failed check is counted once.

Statistics. For pass rates I give 95% Wilson intervals [15]. To compare the same 30 answers scored in two ways, I used the exact McNemar test, which looks only at the tasks where the two results disagree [16]. To compare two different runs, I used Fisher's exact test.

Results

Scores as first recorded. Figure 1 shows the pass counts in the first scoring and after the scoring was corrected. As first recorded, the results looked very different from each other. Qwen

passed 24 of 30 tasks at the default temperature (run A) and 29 at temperature 0 (run C). gpt-oss-120b passed 29 at the default temperature and 29 at temperature 0. North Mini Code passed only 4 tasks. If these numbers were taken at face value, one would conclude that North Mini Code is much weaker than the other two, and that Qwen became much better at temperature 0.

Even so, with only 30 tasks, a gap like 24 against 29 is not strong evidence. The 95% interval for 24 of 30 is 62.7% to 90.5%, and for 29 of 30 it is 83.3% to 99.4%. Fisher's exact test for 24 against 29 gives p = 0.10.

The same answers, scored in two ways. Table 2 shows what happened when the saved answers were scored again. The 30 answers of run A passed 24 tasks when each answer was run alone and 30 tasks when the prompt was placed in front. All six changes went in the same direction, and none of the answers that passed alone failed with the prompt. For these six discordant tasks the exact McNemar test gives p = 0.031. For run B the count went from 29 to 30, and for run C from 29 to 30. Nothing in the code was changed; only the scoring was.

Table 2. Number of tasks passed out of 30 in each way of scoring.

Run	Answer alone	Prompt + answer, as saved	After all corrections
A. Qwen, default	24	30	30
B. gpt-oss, default	29	30	30
C. Qwen, temp. 0	29	30	30
D. gpt-oss, temp. 0	not scored	29	29
E. North Mini Code, temp. 0	not scored	4	29

For run E, "after all corrections" means after removing the markdown fences, re-sending two answers that had been cut off, and repeating the check for HumanEval/15.

In run E the first check was already done with the prompt in front, but 26 of the 30 saved answers were wrapped in markdown fences, even though the prompt asked for no markdown. The reply for HumanEval/10 also had fences, but I had removed them by hand when I saved that answer again. Python cannot run a file that starts with three backticks, so all 26 failed without running a single test. After the fences were removed, 26 tasks passed. Two of the remaining failures, HumanEval/1 and HumanEval/21, were answers that had stopped in the middle at the 2,000-token limit; with a limit of 5,000 tokens both passed. One more, HumanEval/15, failed one of its 137 tests on the first check and passed all 137 when the same code was run again. The code did not change, so the most likely reason is the time limit of the test.

Where the failures came from. Across the five runs there were 35 failed checks. Figure 2 shows what caused them, and Table 3 lists the tasks.

Eight failures were caused by a name that the answer used but did not define. In six cases, five in run A and one in run B, the answer used the type List in the function signature but did not include the line "from typing import List". Python evaluates the signature when it defines the function, so the program stopped with "NameError: name 'List' is not defined" before any test was run. This is why these tasks show zero passed tests and not a partly working function. In the other two cases, runs A and C, the answer to HumanEval/10 called is_palindrome, the helper function that is given in the prompt. In a sense the model did what the task asked: it used the function it was given.

The other failures were 22 answers with markdown fences in run E, two answers cut off at the token limit and one check that failed once and passed when repeated. Only two failures were caused by wrong logic.

Table 3. All failed checks in the five runs.

Run	Task	What went wrong	Fixed by
A	5, 9, 17, 22, 26	"from typing import List" missing, NameError	putting the prompt in front
A, C	10	helper is_palindrome is in the prompt, not in the answer	putting the prompt in front
B	28	"from typing import List" missing, NameError	putting the prompt in front
D	24	wrong logic in largest_divisor	not fixed
E	22 tasks	code wrapped in markdown fences	removing the fences
E	1, 21	answer cut off at 2,000 tokens	new request with 5,000 tokens
E	15	failed 1 of 137 tests on the first check, most likely the time limit	running the check again
E	5	fences, then wrong logic in intersperse	not fixed

The two real errors. In run D, gpt-oss-120b solved HumanEval/24, which asks for the largest divisor of n that is smaller than n, by starting at the square root of n and going down. The first divisor found this way is the largest divisor that is not bigger than the square root, and the function returned n divided by it. For 100 this gives 10, while the correct answer is 50. The code's own comment says it looks for "the smallest factor", but the loop does not do that. The function passed 4 of 5 original tests, so the error shows up only for some inputs. In run B, at the default temperature, the same model solved this task correctly.

In run E, North Mini Code solved HumanEval/5, which asks to put a delimiter between every two neighbouring numbers of a list. Its code paired every number with the delimiter, so the delimiter also ended up after the last number: for [1, 2, 3] and 4 it gives [1, 4, 2, 4, 3, 4] instead of [1, 4, 2, 4, 3]. It passed 7 of 156 tests, only those where the list was empty.

Models and temperature after the correction. With the same scoring for every run, the counts were 30, 30 and 30 for runs A, B and C, and 29 and 29 for runs D and E. The differences between models are at most one task, and Fisher's exact test for 30 against 29 gives $p = 1.0$. For Qwen, temperature 0 and the default temperature gave the same result, 30 of 30. For gpt-oss-120b, the default run passed 30 and the temperature 0 run passed 29, and the one failure was the logic error described above.

Temperature did matter in one respect. When the answers were scored alone, Qwen's default-temperature answers left out the typing import in 5 of the 21 tasks that needed it, while none of the failures of its temperature 0 answers were caused by a missing import. So the difference between 24 and 29 came from whether the model copied the import line into its answer, not from whether it solved the task. With one run for each setting I cannot tell whether this was an effect of temperature or chance.

Discussion. The main result is that, in this small study, most of the failures were in the way the code was handled, not in the code. Of 35 failed checks, 30 disappeared when the answers were taken out of the reply and scored in the same way for every run, and 3 more when

the two cut-off answers were requested again and the unstable check was repeated. What was left, two logic errors in 150 answers, is too little to rank the three models.

This matters for how the first numbers would be read. The gap between Qwen's 24 and 29 looked like an effect of temperature, and North Mini Code's 4 of 30 looked like a weak model. In both cases the difference came from the pipeline and not from the models. This also explains a mistake we made during the project. In the results table for run C, none of the six tasks that Qwen had failed in run A failed again, and we took this as a sign that the model gives different results from run to run. Only after scoring the saved answers of run A again did it become clear that the scoring had also changed between the runs, and that five of the six failures had the same cause, a missing import line. Checking the saved answers, and not only the pass counts, would have shown this at once.

The results fit with what larger studies have found. Biderman and co-authors warn that evaluation details, which are often not reported, can change results [5], and Sclar and co-authors showed how much formatting alone can move a score [6]. Here the "formatting" was on the side of the answer: an import line, a helper function and markdown fences. The EvalPlus tool itself avoids some of these problems. When it generates answers from chat models, it asks for "a self-contained Python script", which would tell the model to include the import line [4]. My prompt asked only to return the code, so an answer without the import line was a reasonable reading of the instruction.

North Mini Code ignored the instruction not to use markdown in 27 of its 30 replies. gpt-oss-120b followed it in run D, where the whole reply was saved. For Qwen it is not possible to say, because in runs A and C the fences were removed when the code was saved. A pipeline that takes the code out of a fenced block, as in runs A, B and C, handles this without problems. A pipeline that saves the whole reply, as in runs D and E, depends on each model obeying the instruction, and the models did not all behave the same way.

A second point is about the token limit. North Mini Code uses part of its token budget for reasoning, and at 2,000 tokens it sometimes ran out before writing the answer. Such cases look like failures of the model, although the cause is a setting that suits one model better than another. Using the same limit for all models does not make a comparison fair when the models spend tokens differently.

Finally, the results agree with the advice to report uncertainty [9, 10]. With 30 tasks, even a result of 30 out of 30 has a 95% interval from 88.6% to 100%. Two models that differ by one task cannot be told apart with this number of tasks, and differences between runs such as those reported by Ouyang and co-authors [7] could not be measured with one run per setting.

Conclusion, limitations and recommendations. I tested three open language models on the first 30 tasks of HumanEval+ and looked at every failure. In the first scoring, the models seemed to differ a lot, and one model seemed to depend strongly on temperature. After the answers were scored in the same way, all three passed 29 or 30 tasks. Only 2 of 35 failed checks were caused by wrong logic; the rest came from how the code was taken out of the reply, how it was run, and the token limit. For a small benchmark, these choices can change the result more than the choice of model does.

Limitations

Few tasks. Thirty tasks, taken from the start of HumanEval+ and not at random, are too few to show small differences between models.

One run per setting. Each model was run once at each temperature, so run-to-run variation could not be measured, and the effect of temperature cannot be separated from chance.

Lost answer files. The answers of runs C, D and E were not saved outside the Colab session. For these runs I rely on the scores printed in the notebook and could not re-score them myself.

Settings changed between runs. Besides temperature, the way the code was saved and scored changed between the runs. This is the main subject of the paper, but it also means the runs are not a clean experiment on temperature.

Model versions. The models were used through Groq and OpenRouter in September 2026. Providers can update a model behind the same name, so the results may not repeat exactly later.

Possible training on the benchmark. HumanEval has been public since 2021 and may be in the training data of these models, which could make the tasks easier for them than new tasks would be.

Recommendations. For anyone who runs a small benchmark of this kind: keep every answer file, not only the scores; decide in advance whether answers are scored as complete programs or together with the prompt, and use the same choice for every model and every run; take the code out of the reply in the same way for all models; check that no answer was cut off at the token limit; and look at the error of every failed task before counting it as a failure of the model.

For future work, the same tasks should be run several times for each model and temperature, which would show how much the pass count changes between runs. A larger and random set of tasks from HumanEval+ or MBPP+ would make the comparison between models more informative.

Data availability. The HumanEval+ dataset and the EvalPlus package are publicly available [4]. The table of results for every task and run, the 60 saved answers of runs A and B, and the script used to score them again are available from the author on request.

AI-use disclosure. The author wrote the Colab notebook with help from ChatGPT, ran the models, and recorded the results. The re-scoring of the saved answers and the figures were done in Python with the help of an AI assistant, and generative AI was also used for language editing and structural revision of the text. No data, results or citations were generated by AI: every pass or fail result comes from running the models' code against the EvalPlus tests, all numbers were checked against the notebook output and the re-scoring, and all cited sources were checked against the originals. The author has reviewed the final manuscript and takes full responsibility for it.

Acknowledgements. I thank my supervisor, Ainabek Onayev, for supervising this project.

Conflicts of interest. The author declares no conflicts of interest.

References:

[1] Sarkar, A., & Drosos, I. (2025). Vibe coding: Programming through conversation with artificial intelligence. arXiv preprint arXiv:2506.23253. https://doi.org/10.48550/arXiv.2506.23253

[2] Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. de O., Kaplan, J., ... Zaremba, W. (2021). Evaluating large language models trained on code. arXiv preprint arXiv:2107.03374. https://doi.org/10.48550/arXiv.2107.03374

[3] Austin, J., Odena, A., Nye, M., Bosma, M., Michalewski, H., Dohan, D., ... Sutton, C. (2021). Program synthesis with large language models. arXiv preprint arXiv:2108.07732. https://doi.org/10.48550/arXiv.2108.07732

[4] Liu, J., Xia, C. S., Wang, Y., & Zhang, L. (2023). Is your code generated by ChatGPT really correct? Rigorous evaluation of large language models for code generation. In Advances in Neural Information Processing Systems 36 (NeurIPS 2023) (pp. 21558-21572). https://doi.org/10.52202/075280-0943

[5] Biderman, S., Schoelkopf, H., Sutawika, L., Gao, L., Tow, J., Abbasi, B., ... Zou, A. (2024). Lessons from the trenches on reproducible evaluation of language models. arXiv preprint arXiv:2405.14782. https://doi.org/10.48550/arXiv.2405.14782

[6] Sclar, M., Choi, Y., Tsvetkov, Y., & Suhr, A. (2023). Quantifying language models' sensitivity to spurious features in prompt design or: How I learned to start worrying about prompt formatting. arXiv preprint arXiv:2310.11324. https://doi.org/10.48550/arXiv.2310.11324

[7] Ouyang, S., Zhang, J. M., Harman, M., & Wang, M. (2025). An empirical study of the non-determinism of ChatGPT in code generation. ACM Transactions on Software Engineering and Methodology, 34(2), 1-28. https://doi.org/10.1145/3697010

[8] Atil, B., Aykent, S., Chittams, A., Fu, L., Passonneau, R. J., Radcliffe, E., ... Baldwin, B. (2024). Non-determinism of "deterministic" LLM settings. arXiv preprint arXiv:2408.04667. https://doi.org/10.48550/arXiv.2408.04667

[9] Miller, E. (2024). Adding error bars to evals: A statistical approach to language model evaluations. arXiv preprint arXiv:2411.00640. https://doi.org/10.48550/arXiv.2411.00640

[10] Blackwell, R. E., Barry, J., & Cohn, A. G. (2024). Towards reproducible LLM evaluation: Quantifying uncertainty in LLM benchmark scores. arXiv preprint arXiv:2410.03492. https://doi.org/10.48550/arXiv.2410.03492

[11] OpenRouter. (n.d.). Qwen: Qwen3.8 27B [model page]. Retrieved September 26, 2026, from https://openrouter.ai/qwen/qwen3.8-27b

[12] OpenAI. (2025). gpt-oss-120b & gpt-oss-20b model card. arXiv preprint arXiv:2508.10925. https://doi.org/10.48550/arXiv.2508.10925

[13] OpenRouter. (n.d.). Cohere: North Mini Code [model page]. Retrieved September 26, 2026, from https://openrouter.ai/cohere/north-mini-code

[14] Groq. (n.d.). API reference: Chat completions. Retrieved September 26, 2026, from https://console.groq.com/docs/api-reference

[15] Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association, 22(158), 209-212. https://doi.org/10.1080/01621459.1927.10502953

[16] McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika, 12(2), 153-157. https://doi.org/10.1007/BF02295996

Figures

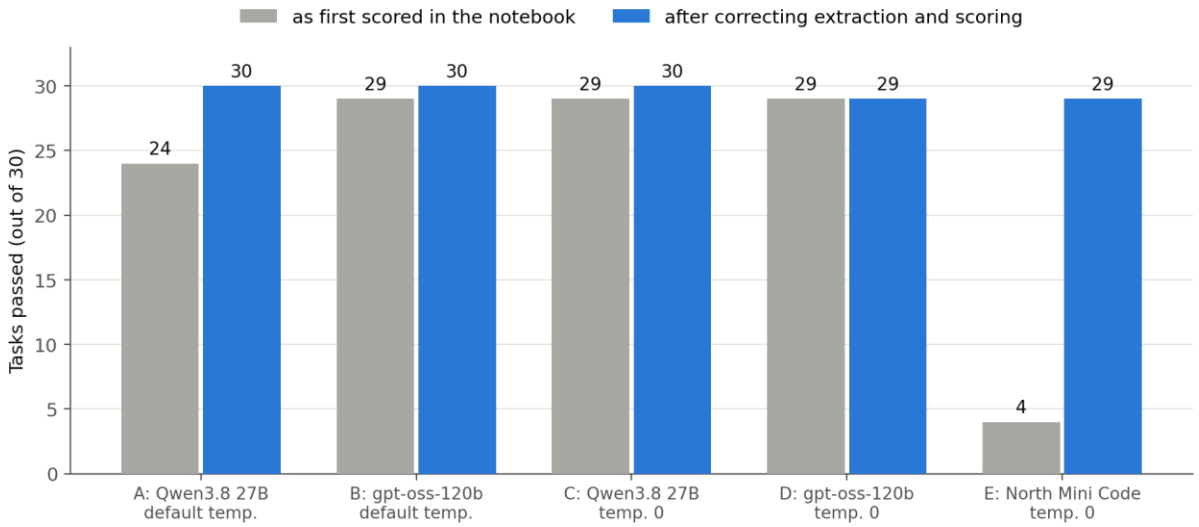


Figure 1. Tasks passed out of 30 in each run, as first scored in the notebook and after the code extraction and scoring were made the same for all runs.

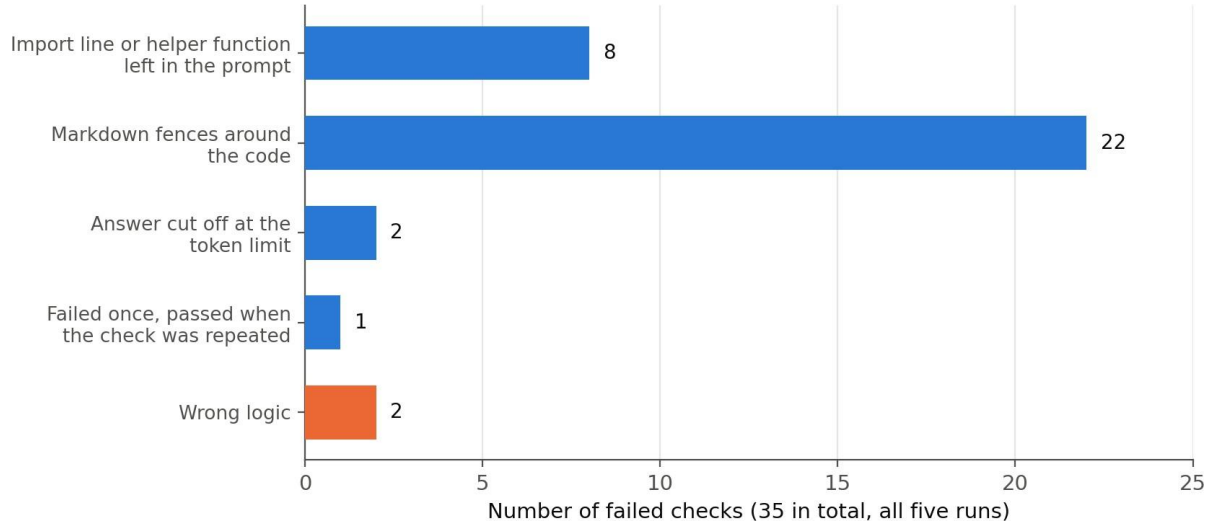


Figure 2. The 35 failed checks in all five runs, grouped by cause. Each check is counted once, under the reason it finally failed for.

UDC 629.783

Nurzhan Sanzhar

Grade 8 student
National School of Physics and Mathematics Almaty (NSPM)
(Almaty, Kazakhstan)
Scientific supervisor: Sydyk Assyltay
2nd-year student,
New York University Abu Dhabi
(Abu Dhabi, UAE)

A FEW SHELLS, A FEW OPERATORS: WHERE LOW EARTH ORBIT IS CROWDED IN AN ALTITUDE-INCLINATION MAP OF 16,620 ACTIVE SATELLITES

Abstract: Orbital congestion is usually described by counting satellites at each altitude. Altitude alone, however, cannot show how the population is organised, because satellites at one altitude may fly in very different orbital planes. Using the CelesTrak catalogue of active satellites retrieved on 28 September 2026 (16,620 objects), we computed each satellite's mean altitude from its mean motion with Kepler's third law and mapped the population on two axes at once, altitude in 50 km bands and inclination in 5 degree bands. Crowding is extremely concentrated. 70.3% of active satellites fly between 450 and 600 km, and the single band 450-500 km holds 9,899 satellites (59.6% of the catalogue), 93.7% of them Starlink. The three densest altitude-inclination cells are all in this band: 50-55 degrees (4,425 satellites, 99.7% Starlink), 40-45 degrees (3,266, 99.4% Starlink) and 95-100 degrees (1,886, 73.7% Starlink). Most of the next-largest concentrations are also single-operator shells: Amazon Kuiper holds 369 of 612 satellites at 600-650 km, OneWeb 648 of 720 at 1,150-1,250 km, and Qianfan 198 of 244 at 1,050-1,100 km. Without these four constellations, no 50 km band holds more than 646 active satellites. The snapshot was taken while SpaceX was moving about 4,400 Starlink satellites from roughly 550 km to 480 km during 2026, so the location of the main peak is itself a recent and changing decision. In the current catalogue, crowding in low Earth orbit is a property of a few shells built by a few operators, and it should be described on altitude and inclination together, with the date of the catalogue.

Keywords: orbital congestion; low Earth orbit; satellite constellations; Starlink; Kepler's third law; CelesTrak

Introduction. The number of active satellites has grown sharply in recent years, driven mainly by large communication constellations in low Earth orbit (LEO) [1, 2]. Concern about this growth goes back to the observation that collisions between orbiting objects can produce fragments which cause further collisions, a cascade first described by Kessler and Cour-Palais [3]. How crowded a region of space is therefore matters beyond the number of satellites in it.

Most popular descriptions of congestion count satellites by altitude. Large constellations, however, are built as shells: many satellites at one altitude and one inclination, spread over several orbital planes [4]. Two populations at the same altitude but at different inclinations are different shells, often run by different operators, while one operator's shell appears as a single dense spot when altitude and inclination are plotted together. A two-dimensional map therefore

shows how the population is organised and who fills each part of it, which a one-dimensional altitude count cannot. It does not, by itself, show where satellites approach each other; that requires conjunction analysis, which is outside the scope of this study.

The timing of this study also matters. In January 2026 SpaceX announced that about 4,400 Starlink satellites would be lowered from roughly 550 km to about 480 km over the course of 2026, stating that debris and planned constellations are less dense below 500 km [5]. A catalogue taken in September 2026 therefore captures the constellation part-way through a deliberate move.

This study asks a narrow question that can be answered from public data: which combinations of altitude and inclination are crowded in the current catalogue of active satellites, and how much of that crowding belongs to single operators?

Materials and Methods. Data. We used the CelesTrak "active" group of general perturbations (GP) element sets, retrieved on 28 September 2026 at 01:09 UTC from https://celestrak.org/NORAD/elements/gp.php?GROUP=active&FORMAT=csv [6]. The file lists 16,620 objects. Their element-set epochs range from 31 August to 27 September 2026, with one predicted epoch dated 29 September. For each object the file gives the name, the international designator (which contains the launch year), mean motion (revolutions per day), eccentricity and inclination. All 16,620 records were complete and none were excluded. "Active" here follows CelesTrak's classification of operational satellites; a few long-lived passive objects, such as the calibration sphere CALSPHERE 1 launched in 1964, are included in that list.

Altitude. For each satellite, the orbital period is T = 86,400 / n seconds, where n is the mean motion in revolutions per day. Kepler's third law gives the semi-major axis a = (μT² / 4π²)^(1/3), with μ = 398,600.4418 km³/s² [7]. Mean altitude is a minus 6,378.137 km, Earth's equatorial radius. Two simplifications should be noted. First, GP mean motion is a mean element defined for the SGP4 model, and converting it directly ignores the correction for Earth's oblateness; for low orbits this changes altitude by about one to two kilometres, which can move a satellite across a band edge but does not change the pattern. Second, altitude is measured above the equatorial radius, so over the poles a satellite is about 21 km higher above the surface than stated. Because 16,570 of the 16,620 objects have eccentricity below 0.1, mean altitude is a good description of where almost all of them fly.

Bands and operators. Satellites were grouped into altitude bands of 50 km (band k covers 50k to 50k + 50 km) and inclination bands of 5 degrees. Operators of the four largest constellations were identified from object names (names containing STARLINK, ONEWEB, KUIPER or QIANFAN). Identification by name is reliable for these constellations but not for every smaller satellite, which is why all other objects are grouped as "other".

Launch year. The launch year was read from the first four characters of the international designator. Because the catalogue contains only satellites that are still active, counts by launch year show how many satellites from each year are in a band today. They do not show when a satellite entered that band, which matters here because many Starlink satellites were launched to higher orbits and later lowered.

Results. Altitude alone. Figure 1 shows the population below 2,000 km (15,806 satellites) by altitude and constellation, and Table 1 lists the most crowded 50 km bands. The band 450-500 km contains 9,899 satellites, 59.6% of all active satellites, and 93.7% of them are Starlink. Together, the three bands from 450 to 600 km hold 11,681 satellites, 70.3% of the catalogue. Other bands are also dominated by one constellation: 600-650 km by Amazon Kuiper (369 of

612), 1,050-1,100 km by Qianfan (198 of 244), and 1,150-1,250 km by OneWeb (648 of 720). The geostationary band at 35,750-35,800 km holds 554 satellites from many operators.

Table 1 - The ten most crowded 50 km altitude bands, with the share of the largest constellation in each

Band, km	Satellites	Largest constellation	Its satellites	Its share
450-500	9,899	Starlink	9,277	93.7%
500-550	944	Starlink	294	31.1%
550-600	838	Starlink	337	40.2%
350-400	817	Starlink	709	86.8%
600-650	612	Amazon Kuiper	369	60.3%
35,750-35,800	554	none	0	0.0%
400-450	480	Starlink	250	52.1%
1,150-1,200	383	OneWeb	315	82.2%
1,200-1,250	337	OneWeb	333	98.8%
1,050-1,100	244	Qianfan	198	81.1%

Altitude and inclination together. Figure 2 shows the population on both axes. Crowding is concentrated in a few cells rather than spread across the band. The three densest cells are all at 450-500 km: inclination 50-55 degrees (4,425 satellites, 4,410 Starlink), 40-45 degrees (3,266, 3,246 Starlink) and 95-100 degrees (1,886, 1,390 Starlink). No other cell below 2,000 km holds more than 550 satellites. The next dense cells are also single-operator shells: 600-650 km at 50-55 degrees (381 satellites, 369 Kuiper), 1,050-1,100 km at 85-90 degrees (Qianfan) and 1,150-1,250 km at 85-90 degrees (OneWeb). The 95-100 degree cells correspond to the sun-synchronous corridor used by Earth-observation satellites; across all altitudes below 2,000 km, the 96-99 degree range holds 3,615 satellites, of which 1,530 are Starlink. Sixteen objects fly on retrograde orbits with inclinations above 100 degrees, up to 142 degrees; Figure 2 extends to 145 degrees to include them.

Removing the large constellations. When Starlink satellites are excluded, the most crowded bands are 500-550 km (650 satellites), 450-500 km (622), 600-650 km (612) and the geostationary band (554). When all four large constellations are excluded, the largest band holds 646 satellites (500-550 km), and the 600-650 km band drops from 612 to 243. The largest band without Starlink is about fifteen times smaller than the 450-500 km band with it.

Launch year of satellites now in the main band. Of the 9,899 satellites now in the 450-500 km band, 118 were launched before 2021 (the oldest in 1990), 491 in 2021, 1,323 in 2022, 1,547 in 2023, 1,568 in 2024, 2,943 in 2025 and 1,909 in 2026 up to 27 September (Figure 3). Because many Starlink satellites launched in 2021-2025 first operated near 550 km and were lowered in 2026 [5], these counts describe the age of the satellites in the band, not when the band filled.

Discussion. The result reframes what "congested low Earth orbit" means in the current catalogue. At the level of the whole population, crowding is not a general property of LEO. It is a property of a small number of shells, each defined jointly by altitude and inclination and each filled mostly by one operator: one very large Starlink shell system at 450-500 km, and smaller shells of Kuiper, Qianfan and OneWeb higher up. Outside these shells the distribution of active satellites is much flatter.

This matters for how congestion is measured. An altitude histogram alone shows one very tall peak at 450-500 km, but it cannot show that most satellites in the peak sit in two inclination bands, 40-45 and 50-55 degrees, arranged as organised shells. The altitude-inclination map shows this directly and makes clear which operator is responsible for each dense region. It also shows that the sun-synchronous corridor at 95-100 degrees is the one dense cell shared by a large constellation and many independent operators, which makes it the natural place to begin a study of crossing traffic.

The map also depends strongly on the date of the catalogue. The peak at 450-500 km is the result of SpaceX's 2026 decision to lower thousands of satellites from about 550 km [5], a move the company presented as a safety measure because debris and planned constellations are less dense below 500 km. The same method applied to a catalogue from early 2026 would place the Starlink peak higher. Any statement about where orbit is crowded should therefore carry the date of the catalogue it describes, and repeated snapshots of the same shells over time would turn this map into a record of how operators move.

Limitations. The catalogue lists active satellites only; it does not include debris, rocket bodies or inactive satellites, which contribute strongly to collision risk [2]. Mean altitude from GP mean motion ignores the oblateness correction and is measured above the equatorial radius, so individual altitudes carry an uncertainty of a few kilometres; mean altitude is also not a good description of the 50 objects with eccentricity above 0.1. Operators were identified by name, which is reliable for the four large constellations but not for all other satellites. Counts by launch year describe satellites still active today and do not show when they reached their present altitude. Finally, the study describes where satellites are, not how often they approach each other; conjunction analysis would be the next step.

Conclusion. In the CelesTrak catalogue of 28 September 2026, 70.3% of active satellites fly between 450 and 600 km, and the three densest altitude-inclination cells are all in the 450-500 km band and are filled mostly by Starlink. Most of the other dense regions are also single-operator shells, of Amazon Kuiper, Qianfan and OneWeb, and without these four constellations no altitude band holds more than 646 active satellites. Crowding in the current catalogue is therefore concentrated, dated and largely built by a few operators, and it should be measured on altitude and inclination together.

Data availability. The catalogue is publicly available from CelesTrak at the address given in Materials and Methods. The analysis code is available from the authors on request.

Conflict of interest. The authors declare no conflict of interest.

References:

1. ESA Space Debris Office. ESA's Annual Space Environment Report 2025. European Space Agency, 2025. URL: https://www.esa.int/Space_Safety/Space_Debris/ESA_Space_Environment_Report_2025 (accessed: 27.09.2026).

2. Boley A.C., Byers M. Satellite mega-constellations create risks in Low Earth Orbit, the atmosphere and on Earth // Scientific Reports. 2021. Vol. 11. Art. 10642. DOI: 10.1038/s41598-021-89909-7.

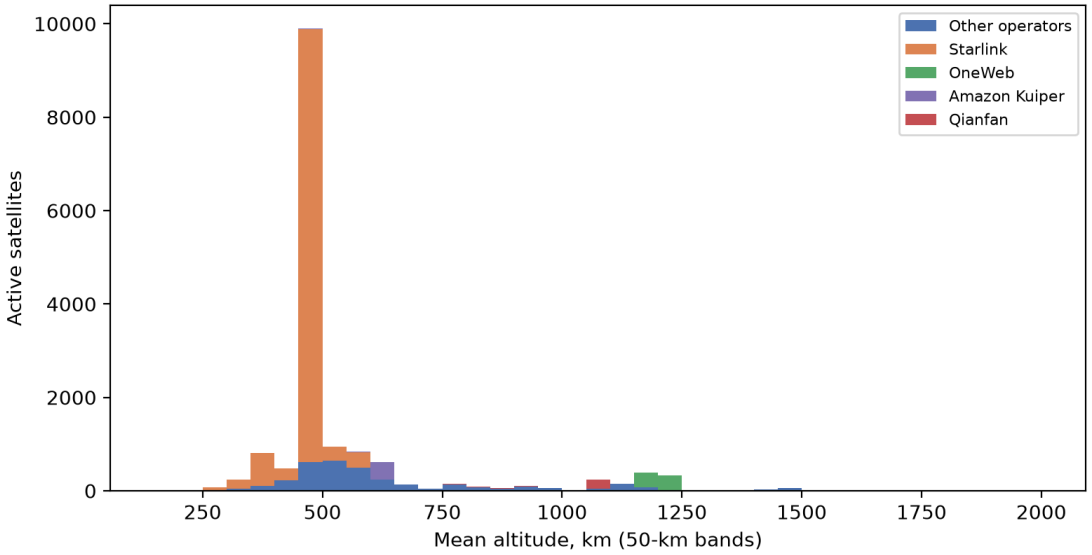
3. Kessler D.J., Cour-Palais B.G. Collision frequency of artificial satellites: The creation of a debris belt // Journal of Geophysical Research. 1978. Vol. 83, No. A6. P. 2637-2646. DOI: 10.1029/JA083iA06p02637.

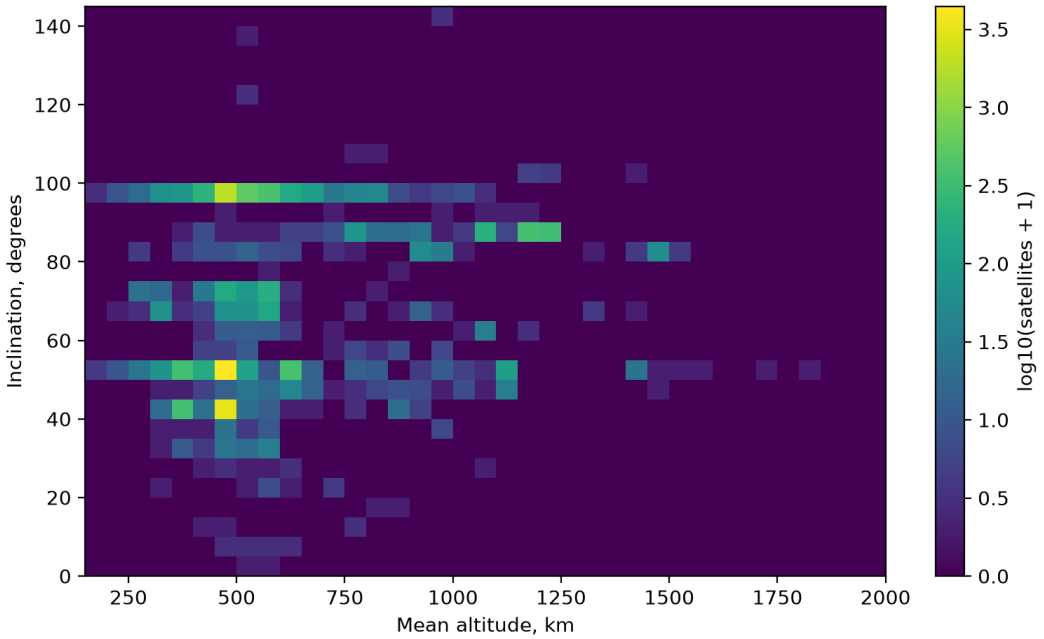
4. McDowell J.C. The Low Earth Orbit Satellite Population and Impacts of the SpaceX Starlink Constellation // The Astrophysical Journal Letters. 2020. Vol. 892, No. 2. Art. L36. DOI: 10.3847/2041-8213/ab8016.

5. Robinson D. Starlink to lower satellite orbits over safety concerns // The Register. 02.01.2026. URL: https://www.theregister.com/2026/01/02/starlink_lower_orbits/ (accessed: 28.09.2026).

6. Kelso T.S. CelesTrak: A New Way to Obtain GP Data. URL: https://celestrak.org/NORAD/documentation/gp-data-formats.php (accessed: 28.09.2026).

7. Vallado D.A. Fundamentals of Astrodynamics and Applications. 4th ed. Hawthorne, CA: Microcosm Press, 2013.





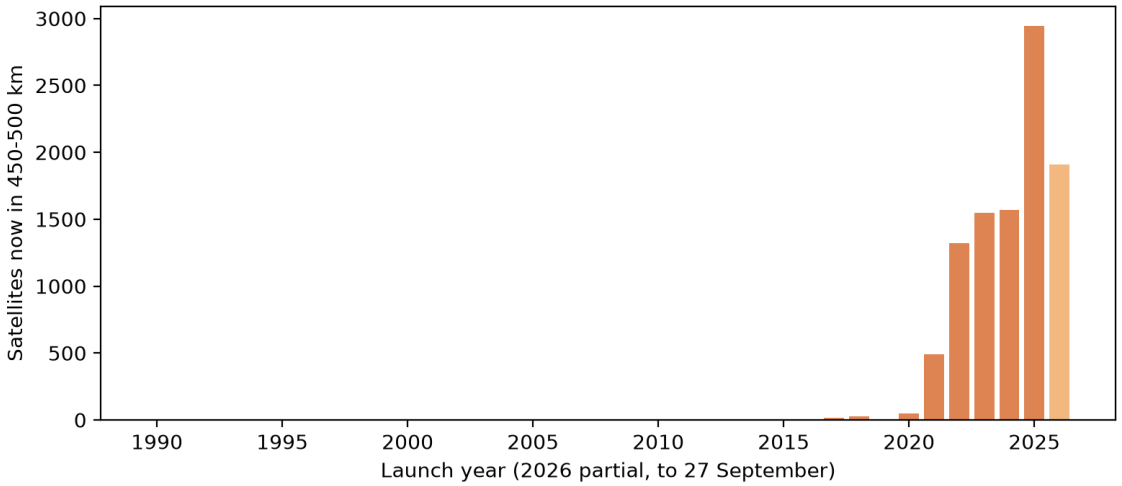


Figure captions:

Figure 1. Active satellites below 2,000 km by mean altitude (50 km bands), coloured by constellation. CelesTrak catalogue, 28 September 2026.

Figure 2. Number of active satellites in each altitude-inclination cell (50 km by 5 degrees), below 2,000 km and up to 145 degrees inclination, on a logarithmic colour scale.

Figure 3. Satellites now in the 450-500 km band by launch year. The 2026 bar covers launches to 27 September only.

UDC 504.3.054:004.85

Askarbek Nurislam
Grade 9 student
Supervisor: Yergashbayev Zhasur
computer science teacher
Astana Bilim-Innovation Lyceum for Boys
(Astana, Kazakhstan)

CAN THE WEATHER TELL WHICH DAYS IN ASTANA WILL BE POLLUTED? CLASSIFYING PM2.5 EXCEEDANCE DAYS WITH MACHINE LEARNING AND CHECKING THE DATA AGAINST A GROUND MONITOR

Abstract: Astana has long cold winters, and part of the city is still heated with coal. In this study I asked whether weather data alone can show in advance on which days the daily mean concentration of fine particles (PM2.5) in Astana will be above the World Health Organization guideline of 15 μg/m³. I used free hourly PM2.5 estimates from the Copernicus Atmosphere Monitoring Service (CAMS) and ERA5 weather data for 1,490 days between 4 August 2022 and 1 September 2026. Only 141 days (9.5%) were above the guideline, so a model that always answers "normal" is right 90.5% of the time and still finds no polluted day. I trained logistic regression, random forest and gradient boosting models on 2022-2024 and tested them once on 2025-2026. With the weather of the same day, the random forest found 64% of the exceedance days with 55% of its alarms correct (F1 = 0.59, 95% CI 0.47-0.68), better than the rule "tomorrow will be like today" (F1 = 0.42). When the model could use only the weather of the previous day, F1 fell to 0.32, below that simple rule. The model missed all ten summer exceedances and most days with a lot of dust in the air. Finally, I compared CAMS with the US Embassy monitor in Astana on 613 days. The monitor showed about twice as much PM2.5 (17.4 against 8.4 μg/m³ on average) and exceedances on 39% of days instead of 9%, and only the monitor showed higher pollution in the heating season. Weather predicted the monitor's exceedance days at least as well as the CAMS ones. My conclusion is that weather contains real information about polluted days in Astana, but CAMS shows a much cleaner city than the ground monitor, so models trained only on CAMS should be used with care.

Keywords: PM2.5; air quality; Astana; machine learning; imbalanced classification; CAMS; ground-based monitoring

1. Introduction

I live in Astana, and every winter there are evenings when the air smells of smoke and the city looks grey. Part of the city is still heated with coal, and the heating season lasts about half a year. Air pollution in Astana and its monitoring system have been studied before [1]. Studies describe coal heating and thermal power plants close to the city as important sources of particulate pollution [2], and a recent sampling study found soot and coal fly ash in Astana's particulate matter and PM2.5 concentrations up to 534 μg/m³ in individual samples [3].

PM2.5 means particles with a diameter of 2.5 micrometres or less. They go deep into the lungs, and in 2021 the World Health Organization lowered its guideline for the 24-hour mean of PM2.5 to 15 μg/m³ [4]. For a person living in the city, the exact number is less useful than a simple answer to the question "will tomorrow be a bad day?". This is why I decided to treat the

problem as a classification problem with two classes, "normal" and "exceedance", instead of predicting the exact concentration.

Predicting PM2.5 with machine learning is a common topic. Studies from Iran used meteorological data and several algorithms to predict PM2.5 levels or concentrations [5, 6], and a study in Frontiers in Big Data classified PM2.5 into several groups and used oversampling because the classes were imbalanced [7]. Imbalance is a known problem in air quality classification [8]. When one class is rare, accuracy is misleading, and precision-recall curves and the Matthews correlation coefficient describe a model better [9, 10].

Most such studies use measurements from ground stations. For Astana I could not find a long and complete public record of measurements, so my plan was to use the free PM2.5 estimates of the Copernicus Atmosphere Monitoring Service (CAMS) through the Open-Meteo service [11, 12]. CAMS values are not measurements. They come from an atmospheric model that combines emissions, meteorology and satellite observations [13]. Studies that compared CAMS with ground stations over China [14], around the world using the OpenAQ platform [15] and in Hungary [16] show that CAMS PM2.5 should be checked before use, and in Hungary the raw CAMS data systematically underestimated the measured values [16]. There is also a risk of circular reasoning. CAMS itself uses weather, so a model that predicts CAMS PM2.5 from weather may only learn how CAMS works, not how pollution in the real city behaves.

I did not find a public study for Astana that treats the problem as classification of exceedance days, compares the model with honest baselines on imbalanced data, and checks the data source against a ground monitor. My search was in English and Russian, and I do not claim it was complete. In this study I asked four questions:

• How much better than simple baselines can weather-based models find the days when PM2.5 in Astana is above 15 μg/m³?

• Which weather features are the most important, and does this agree with what is known about how pollution accumulates and disperses?

• What do the days on which the model is wrong have in common?

• How well do the CAMS values agree with a ground monitor in Astana?

Before the analysis I wrote four hypotheses. H1: a model with only weather features will beat the baseline "always normal" in F1 for the exceedance class, but not necessarily in accuracy. H2: the most important features will be those connected to dispersion, such as low wind speed, wind direction and stagnant conditions. H3: missed exceedances will mostly be days when the pollution was caused by something other than the weather, which CAMS smooths out. H4: CAMS will agree with the ground monitor only weakly, especially at low concentrations.

2. Data and methods

PM2.5. I downloaded hourly PM2.5 for the point 51.18° N, 71.45° E from the Open-Meteo Air Quality API [12]. For Astana this API gives the CAMS global atmospheric composition forecast with a resolution of 0.4° (about 45 km), which is available from August 2022. This explains why my data start on 4 August 2022. At first I thought the data were a reanalysis, but the documentation shows that for Central Asia they are archived forecasts, so I call them "CAMS estimates". All hours are in the fixed time zone GMT+5, and each day's mean is the average of 24 hourly values. The first day, 4 August 2022, has 19 hours and I kept it, so the full set has 1,490 days up to 1 September 2026.

Weather. Hourly weather came from the ERA5 reanalysis [17] through the Open-Meteo Historical Weather API [11], for the nearest ERA5 grid point (51.25° N, 71.5° E). I used air

temperature, relative humidity, surface pressure, wind speed and direction at 10 m, precipitation and cloud cover. I requested only ERA5, so that all years come from one model. Open-Meteo also gives the boundary layer height, which is physically very important for pollution, but in this archive it had no values from January to June 2024, so I did not use it in the models.

Ground monitor. The US Embassy in Astana had a PM2.5 monitor whose data were published through the AirNow programme. I could not find this monitor in the OpenAQ search. Its records are in the open OpenAQ data archive as location 7094, "US Diplomatic Post: Astana", and they can be downloaded without registration [18]. Hours with the code -999 were treated as missing. The time stamps in the files change from UTC+6 to UTC+5 in March 2024, when Kazakhstan moved to a single time zone, so I converted every hour to UTC first and then to GMT+5. A station day was used only if it had at least 18 valid hours.

Target and features. A day was an exceedance (y = 1) if its daily mean CAMS PM2.5 was above 15 μg/m³. From the hourly weather I calculated daily mean, minimum and maximum temperature and their range, mean humidity, mean pressure and its change from the previous day, mean and maximum wind speed, the number of calm hours (wind below 2 m/s), wind direction as the sine and cosine of the vector-mean direction, total precipitation and mean cloud cover. I also added the previous day's temperature and wind speed, the change in temperature, the month (as sine and cosine) and a heating season flag (October-March).

Three modes. In mode A the model gets the weather of the same day. This is the setting of my research question, and it is also the most optimistic one, because real forecasters know tomorrow's weather only from a forecast. In mode B the model also gets the previous day's PM2.5, which makes the task easier because pollution changes slowly. I added mode P after the first results: the model gets only the weather of the previous day and the calendar, which is what is really known on the evening before.

Splitting the data. Neighbouring days are similar, so a random split would let the model see the test days through their neighbours. I used a chronological split: training on 4 August 2022 - 31 December 2024 (880 days after removing the first day, which has no previous day, 67 exceedances) and testing on 1 January 2025 - 1 September 2026 (609 days, 74 exceedances). Hyperparameters were chosen by expanding-window cross-validation inside the training part, using the area under the precision-recall curve (PR-AUC). The decision threshold was also chosen inside the training part, as the threshold that gave the best F1 on the cross-validation predictions. The test part was used only once.

Models. I used two baselines: "always normal" and the persistence rule "today is an exceedance if yesterday was". The models were logistic regression with class weights (with standardised features), random forest [19] with balanced class weights, and gradient boosting with balanced sample weights. Everything was done in Python with pandas and scikit-learn [20], with a fixed random seed.

Metrics. For the exceedance class I report precision, recall, F1, PR-AUC and the Matthews correlation coefficient (MCC) [10]. PR-AUC is more informative than ROC-AUC when one class is rare [9]. The test part has only 74 exceedance days, so I calculated 95% confidence intervals with a block bootstrap: 2,000 resamples made of blocks of 7 consecutive days, so that the dependence between neighbouring days is kept. To compare two models, I calculated the difference in F1 on the same resamples.

Error analysis. For the main model I divided the test days into found exceedances (TP), missed exceedances (FN), false alarms (FP) and correct normal days (TN), and compared their

weather, season, position in an episode of polluted days and the amount of dust in the CAMS data. I compared TP and FN with the Mann-Whitney test.

3. Results

3.1 What the CAMS data show

The daily mean CAMS PM2.5 for the whole period was 8.73 µg/m³ (median 7.83, standard deviation 5.27). The 90th percentile was 14.8 µg/m³, almost exactly at the threshold, which means that exceedances are rare by definition. The highest daily mean was 49.9 µg/m³ on 28 November 2024, and the highest hourly value was 65.9 µg/m³ on the same day (Figure 1). Only 141 of 1,490 days (9.5%) were above 15 µg/m³, so the "always normal" model has an accuracy of 90.5% on the full data and 87.8% on the test part.

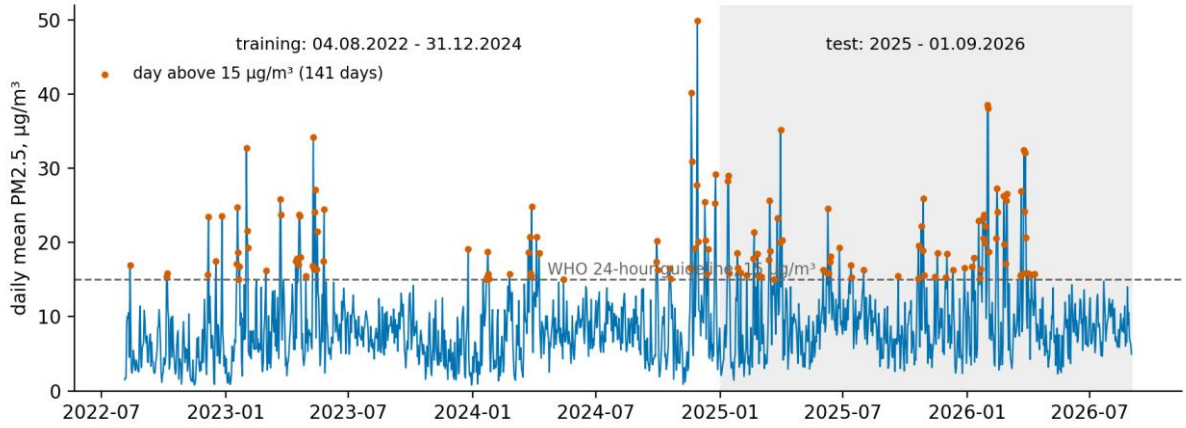


Figure 1. Daily mean CAMS PM2.5 in Astana, 4 August 2022 - 1 September 2026. Orange points are days above the WHO guideline; the grey area is the test period

When I first counted exceedances by calendar year, it looked like the share of exceedance days grew more than twice, from 4.7% in 2022 to 11.8% in 2025 (Table 1). This comparison was wrong, because 2022 starts in August and misses January-March, which in these data are some of the most polluted months. When I compared only August-December, which all years have, 2023 turned out to be the cleanest year (0.7%), then there was a jump in 2024 (11.8%) and a small fall in 2025 (9.2%). So the honest description is a jump between 2023 and 2024 without a clear trend after it.

Table 1. CAMS PM2.5 exceedance days in Astana by year

Year	Days	Exceedance days	Share, %	Mean PM2.5, µg/m³	Aug-Dec share, %
2022	150	7	4.7	5.95	4.7
2023	365	28	7.7	8.24	0.7
2024	366	32	8.7	8.80	11.8
2025	365	43	11.8	9.28	9.2
2026	244	31	12.7	10.22	-

2022 starts on 4 August and 2026 ends on 1 September, so the yearly shares cannot be compared directly; the last column uses only August-December.

In the CAMS data the heating season was almost as clean as the warm season (8.81 against 8.64 µg/m³), and the most polluted month was March, which also had the most dust in the CAMS data (12.4 µg/m³ on average, compared with about 1 µg/m³ in August and

December). I expected the opposite, and section 3.5 shows that the ground monitor disagrees with CAMS here.

3.2 How well the models find exceedance days

Table 2 shows the results on the test period. The "always normal" model has an F1 of zero. The persistence rule found 42% of the exceedance days, with 42% of its alarms correct (F1 = 0.42).

Table 2. Results on the test period (609 days, 74 exceedance days)

Mode and model	Precision	Recall	F1 [95% CI]	PR-AUC [95% CI]	MCC
Always "normal"	-	0.00	0.00	-	0.00
Same as yesterday	0.42	0.42	0.42 [0.30-0.51]	0.46 [0.33-0.58]	0.34
A: logistic regression	0.30	0.82	0.44 [0.34-0.53]	0.51 [0.36-0.66]	0.39
A: random forest	0.55	0.64	0.59 [0.47-0.68]	0.54 [0.39-0.70]	0.53
A: gradient boosting	0.41	0.53	0.46 [0.34-0.56]	0.45 [0.32-0.60]	0.38
B: logistic regression	0.48	0.59	0.53 [0.42-0.62]	0.60 [0.47-0.71]	0.46
B: random forest	0.61	0.61	0.61 [0.49-0.70]	0.69 [0.56-0.79]	0.55
B: gradient boosting	0.53	0.66	0.59 [0.48-0.67]	0.58 [0.43-0.70]	0.53
P: logistic regression	0.23	0.69	0.34 [0.24-0.43]	0.30 [0.20-0.43]	0.24
P: random forest	0.21	0.66	0.32 [0.23-0.41]	0.32 [0.21-0.45]	0.22
P: gradient boosting	0.18	0.78	0.30 [0.22-0.38]	0.23 [0.15-0.33]	0.20

A: weather of the same day; B: A plus yesterday's PM2.5; P: only yesterday's weather and the calendar. Confidence intervals are from a 7-day block bootstrap. For the persistence rule PR-AUC uses yesterday's PM2.5 as the score.

In mode A the random forest was the best model. It found 47 of the 74 exceedance days, gave 39 false alarms and missed 27 exceedances, and its accuracy was 89.2%. This is a little higher than 87.8% for "always normal", but the important difference is in F1: 0.59 against 0. The random forest was also better than the persistence rule (difference in F1 0.17, 95% CI 0.05-0.28) and better than logistic regression (difference 0.15, CI 0.06-0.23). Logistic regression found the most exceedances (82%), but it gave 141 false alarms, so only 30% of its alarms were right. Gradient boosting was not better than logistic regression.

Adding yesterday's PM2.5 (mode B) raised PR-AUC from 0.54 to 0.69 for the random forest, but the F1 improvement was small and not clear (difference 0.02, CI -0.07 to 0.10). For logistic regression the gain from mode B was clearer (difference 0.09, CI 0.003-0.17).

The result that surprised me most was mode P. When the models could use only the weather of the day before, F1 fell to 0.30-0.34, and the random forest was worse than the simple persistence rule (difference -0.10, CI -0.18 to -0.005). Figure 2 shows the same thing with

precision-recall curves: the curve for mode P lies far below the curves for modes A and B. So most of the skill of mode A comes from knowing the weather of the same day.

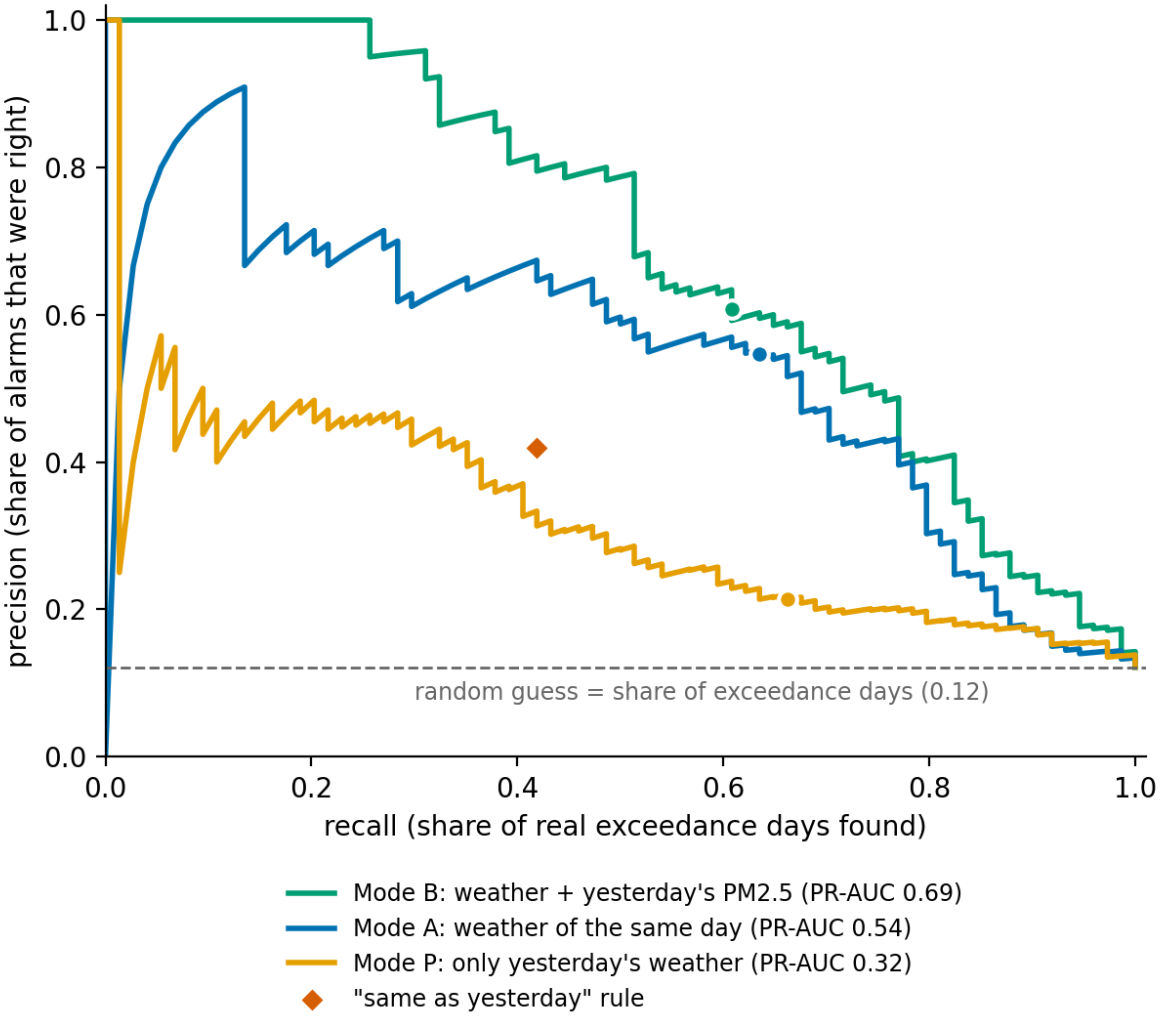


Figure 2. Precision-recall curves of the three random forest models on the test period. Points show the chosen thresholds; the diamond is the "same as yesterday" rule

One more thing I noticed is that cross-validation inside the training part gave much lower PR-AUC values (0.14-0.19 in mode A) than the test. The training years had fewer exceedances, especially 2023 with only one in August-December, so some validation folds had very few positive days. The test result may therefore be a bit lucky, and I do not think the exact numbers would repeat in another year.

3.3 Which features matter

In the random forest (mode A), the features that most reduced PR-AUC when they were shuffled were the maximum wind speed of the day, the previous day's temperature, the minimum temperature and the mean wind speed. Wind direction had only a small effect. In logistic regression, the month, more calm hours, higher pressure and a larger daily temperature range increased the chance of an exceedance, and stronger wind decreased it. Many of these features are correlated with each other, so I read the signs of the coefficients only as a rough picture. In mode B yesterday's PM2.5 was the most important feature, and after it again wind and

temperature. This agrees with the idea that pollution builds up when the air is calm and cold, and it partly supports H2, except that wind direction was not important.

3.4 Where the model makes mistakes

The 27 missed exceedances were different from the 47 found ones. On the missed days the median wind speed was higher (3.4 against 2.3 m/s, $p < 0.001$), there were no calm hours (median 0 against 7), the median temperature was 5.1 °C instead of -11.4 °C, and the ratio of dust to PM10 in the CAMS data was about twice as high (0.21 against 0.10, $p = 0.003$). In other words, the model learned that a polluted day is a cold and calm day, and it missed the polluted days that did not look like that. The position in an episode did not explain the misses: 56% of the missed days and 60% of the found days were the first day of a polluted period.

Three patterns explain most misses. First, many misses were close to the threshold: 63% of the missed days had PM2.5 between 15 and 18 µg/m³, compared with 34% of the found days, so a part of the errors comes from the sharp threshold itself. Second, the model did not find any of the 10 exceedances in June-August 2025. The training years had only one summer exceedance, so the model had nothing to learn from. On several of these June days the CAMS aerosol optical depth was 0.4-0.5, about twice its usual summer value of 0.19, which points to aerosol brought from somewhere else, but from my data I cannot say whether it was dust or smoke. Third, on days when CAMS dust was above 10 µg/m³, the model found only 7 of 18 exceedances, while on the other days it found 40 of 56.

The 39 false alarms were mostly cold, calm days in winter and spring when PM2.5 stayed below the threshold. Their median PM2.5 was 10.7 µg/m³, and 36% of them were above 12 µg/m³, so many false alarms were "almost exceedances".

This means that H3 was only partly right. The missed days were not explained by local sources, because CAMS does not see local sources anyway. They were days when the particles probably came from outside the city, or days very close to the threshold.

3.5 CAMS against the ground monitor

The US Embassy monitor had valid data from 20 October 2022 to 4 March 2025, but with gaps. There were 613 days (24 October 2022 - 3 March 2025) with at least 18 hours of measurements, which is 65% of the CAMS days in this period. Table 3 compares the two sources on these days.

Table 3. CAMS and the US Embassy monitor on the same 613 days

Measure	CAMS	US Embassy monitor
Mean PM2.5, µg/m³	8.4	17.4
Median PM2.5, µg/m³	7.5	12.6
Highest daily mean, µg/m³	49.9	117.8
Share of exceedance days, %	8.8	39.0
Heating season mean (Oct-Mar), µg/m³	7.8	20.4
Warm season mean (Apr-Sep), µg/m³	9.1	13.8

The highest CAMS value is for the whole period; the monitor's highest value was on 6 December 2022. Days are in GMT+5.

The two sources moved together only moderately (Pearson $r = 0.49$, Spearman $\rho = 0.43$). On days when the monitor was below 15 µg/m³ the correlation was only 0.28, which agrees with H4. CAMS was lower on average by 9.0 µg/m³, and the difference was largest in the heating season (Figure 3a). The monitor recorded an exceedance on 239 days, and CAMS also

showed an exceedance on only 50 of them. CAMS showed an exceedance without the monitor on 4 days. The two sources agreed on 68.5% of days, and Cohen's κ was 0.23.

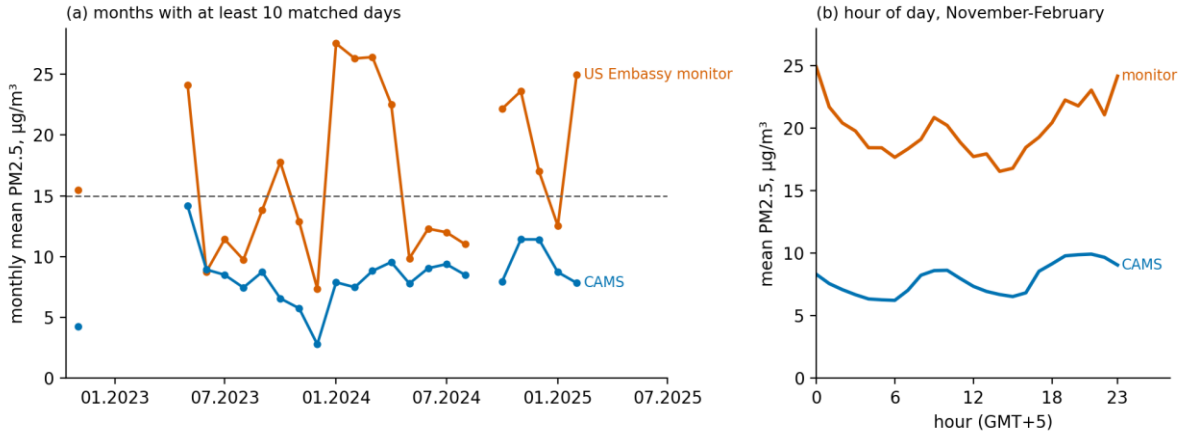


Figure 3. CAMS and the US Embassy monitor: (a) monthly means on matched days; (b) mean hour-of-day profile in November-February

The most important difference is the season. In the CAMS data the heating season was cleaner than the warm season, but at the monitor it was about one and a half times more polluted (20.4 against 13.8 µg/m³), and half of the heating season days were above the guideline. In winter, CAMS shows two peaks, around 9-10 in the morning and 19-21 in the evening. The monitor also has a small morning peak, but its highest values are around midnight, and the night stays polluted (Figure 3b). This fits the picture of houses burning coal in the evening while cold, stable air holds the smoke near the ground.

3.6. Does weather explain the real monitor as well?

To check the circularity problem, I trained the same random forest (mode A) on the 613 matched days twice, once with the CAMS exceedances as the target and once with the monitor's exceedances, using the same expanding-window cross-validation. Because the share of positive days was very different (9% and 39%), I compared ROC-AUC, which does not depend on this share. It was 0.78 for CAMS and 0.82 for the monitor. So the weather predicted the real measurements at least as well as the CAMS estimates.

The relationships were not the same, though. Wind speed and calm hours were related to PM2.5 in both sources in a similar way. Temperature was not: the monitor's PM2.5 was higher on colder days (Spearman $\rho = -0.30$), while CAMS PM2.5 was slightly higher on warmer days ($\rho = 0.15$). Pressure ($\rho = 0.44$ against 0.06) and precipitation ($\rho = -0.42$ against -0.18) were also much more strongly related to the monitor than to CAMS.

4. Discussion

H1 was supported in mode A: the random forest clearly beat both baselines in F1, while its accuracy was only a little higher than that of "always normal". Mode P shows the limit of this result. With only yesterday's weather, the models were worse than the persistence rule, so in a real forecast the quality would depend mostly on the quality of the weather forecast for the next day. My mode A is closer to the case of a perfect weather forecast, and the real result would be somewhere between modes A and P.

H2 was partly supported. Wind speed, calm hours and low temperature were the most important features, but wind direction was not. Maybe the 45 km CAMS cell is too large for the direction of local sources to matter.

H3 was not supported in the form I wrote it. I expected missed days to be caused by local sources that CAMS smooths out, but local sources are missing from the CAMS target as well, so the model could not miss them. The missed days were warm, windy and often dusty days, or days just above 15 μg/m³. This shows that a "weather only" model does not know where particles come from, and that one threshold creates many borderline errors.

H4 was supported, and the comparison with the monitor was the most important part of this work for me. CAMS showed about half of the PM2.5 measured at the US Embassy, and it showed four times fewer exceedance days. The finding that the heating season was not more polluted, which I wrote in my first draft, turned out to come from the CAMS data and not from the real city. A model trained on CAMS learns the rule "exceedances happen on calm days, and dust can add some", while the monitor shows that in Astana cold winter days are polluted much more often. At the same time, the circularity worry was not confirmed: weather explained the monitor's exceedances at least as well as the CAMS ones. I think the right conclusion is that weather data are useful for this task, but the target should be measured PM2.5, and CAMS should be calibrated against stations before it is used for Astana, as was done for Hungary [16].

During the work I also learned two practical lessons. The first is that comparing incomplete years with full years can create a false trend, as in Table 1. The second is that time zones matter. Kazakhstan changed its time zone in 2024, and if I had used the station's local hours without converting them, the hour-of-day profiles would be shifted by one hour for part of the period.

Limitations. The main limitation is that the model's target is a CAMS estimate, which section 3.5 shows to be far below the measured values. The test period has only 74 exceedance days, so the confidence intervals are wide, and the cross-validation results inside the training part were much weaker than the test results. The data cover only four years, and the jump between 2023 and 2024 shows that conditions change from year to year. The comparison uses one monitor at one place in the city, which had data on 65% of the days and stopped publishing on 4 March 2025, so it does not cover the test period. I used observed weather, not weather forecasts. I did not use emissions, traffic or the boundary layer height, and I tested only one threshold (15 μg/m³).

5. Conclusion

Using 1,490 days of CAMS and weather data for Astana, I showed that a random forest using the weather of the same day can find about two thirds of the days with PM2.5 above the WHO guideline, with about half of its alarms correct, which is clearly better than "always normal" and the persistence rule. With only the previous day's weather, the models lost this advantage. The model missed summer exceedances and dusty days, which were rare in the training years. The comparison with the US Embassy monitor showed that CAMS greatly underestimates PM2.5 in Astana, especially in the heating season, and finds only about one in five of the exceedance days that the monitor recorded. For the next step I would like to train the same models on measured PM2.5 from Astana's ground stations, include the boundary layer height and real weather forecasts, and check whether a model trained on measured data can warn people about polluted days one day in advance.

Author contribution and use of AI

The author chose the topic, formulated the research question and hypotheses, wrote the research plan, downloaded the CAMS data, calculated the yearly statistics and the August-December comparison, found the problem with comparing incomplete years, described the hour-of-day pattern and the circularity problem, and collected the first list of literature. The

author's mentor found the US Embassy monitor in the OpenAQ archive, built the models and the comparison with the monitor, made the figures, and helped write and edit the English text. Generative AI tools were used to help with editing and translation. Every number in the paper was calculated by the analysis code from the open data sources and checked by the author's mentor.

References:

1. Kerimray A., Bakdolotov A., Sarbassov Y., Inglezakis V., Poulopoulos S. Air pollution in Astana: analysis of recent trends and air quality monitoring system // Materials Today: Proceedings. 2018. Vol. 5, No. 11. P. 22749-22758. DOI: 10.1016/j.matpr.2018.07.086.

2. Biloshchytskyi A., Neftissov A., Kuchanskyi O., Andrashko Y., Biloshchytska S., Mukhatayev A., Kazambayev I. Fractal analysis of air pollution time series in urban areas in Astana, Republic of Kazakhstan // Urban Science. 2024. Vol. 8, No. 3. Art. 131. DOI: 10.3390/urbansci8030131.

3. Agibayeva A., Kumisbek A., Nauyryzbay A., Avcu E., et al. Towards sustainable air quality in coal-heated cities: a case study from Astana, Kazakhstan // Sustainability. 2025. Vol. 17, No. 22. Art. 10214. DOI: 10.3390/su172210214.

4. World Health Organization. WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Geneva: World Health Organization, 2021. 273 p. ISBN 978-92-4-003422-8. URL: https://www.who.int/publications/i/item/9789240034228 (accessed: 02.10.2026).

5. Mohammadi F., Teiri H., Hajizadeh Y., Abdolahnejad A., Ebrahimi A. Prediction of atmospheric PM2.5 level by machine learning techniques in Isfahan, Iran // Scientific Reports. 2024. Vol. 14. Art. 2109. DOI: 10.1038/s41598-024-52617-z.

6. Makhdoomi A., Sarkhosh M., Ziaei S. PM2.5 concentration prediction using machine learning algorithms: an approach to virtual monitoring stations // Scientific Reports. 2025. Vol. 15. Art. 8076. DOI: 10.1038/s41598-025-92019-3.

7. Saminathan S., Malathy C. Ensemble-based classification approach for PM2.5 concentration forecasting using meteorological data // Frontiers in Big Data. 2023. Vol. 6. Art. 1175259. DOI: 10.3389/fdata.2023.1175259.

8. Ravindiran G., Karthick K., Sivarethinamohan S., Rajamanickam S., et al. A cluster-aware synthetic resampling and machine learning framework for multi-class Air Quality Index classification // Environmental Monitoring and Assessment. 2026. Vol. 198, No. 3. Art. 270. DOI: 10.1007/s10661-026-15121-7.

9. Saito T., Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets // PLOS ONE. 2015. Vol. 10, No. 3. Art. e0118432. DOI: 10.1371/journal.pone.0118432.

10. Chicco D., Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation // BMC Genomics. 2020. Vol. 21. Art. 6. DOI: 10.1186/s12864-019-6413-7.

11. Zippenfenig P. Open-Meteo.com Weather API [software]. Zenodo, 2024. DOI: 10.5281/zenodo.7970649.

12. Open-Meteo. Air Quality API: CAMS global atmospheric composition forecasts, 0.4°, from August 2022. URL: https://open-meteo.com/en/docs/air-quality-api (accessed: 02.10.2026).

13. Rémy S., Kipling Z., Huijnen V., Flemming J., et al. Description and evaluation of the tropospheric aerosol scheme in the Integrated Forecasting System (IFS-AER, cycle 47R1) of ECMWF // Geoscientific Model Development. 2022. Vol. 15, No. 12. P. 4881-4912. DOI: 10.5194/gmd-15-4881-2022.

14. Ali M.A., Bilal M., Wang Y., Nichol J.E., et al. Accuracy assessment of CAMS and MERRA-2 reanalysis PM2.5 and PM10 concentrations over China // Atmospheric Environment. 2022. Vol. 288. Art. 119297. DOI: 10.1016/j.atmosenv.2022.119297.

15. Jin C., Wang Y., Li T., Yuan Q. Global validation and hybrid calibration of CAMS and MERRA-2 PM2.5 reanalysis products based on OpenAQ platform // Atmospheric Environment. 2022. Vol. 274. Art. 118972. DOI: 10.1016/j.atmosenv.2022.118972.

16. Qor-el-aine A., Béres A., Géczi G. Calibration of CAMS PM2.5 data over Hungary: a machine learning approach // Environmental Research Communications. 2024. Vol. 6, No. 7. Art. 075026. DOI: 10.1088/2515-7620/ad6239.

17. Hersbach H., Bell B., Berrisford P., Hirahara S., et al. The ERA5 global reanalysis // Quarterly Journal of the Royal Meteorological Society. 2020. Vol. 146, No. 730. P. 1999-2049. DOI: 10.1002/qj.3803.

18. OpenAQ. Open data archive, location 7094 "US Diplomatic Post: Astana" (AirNow), hourly PM2.5, 2022-2025. URL: https://openaq-data-archive.s3.amazonaws.com (accessed: 02.10.2026).

19. Breiman L. Random forests // Machine Learning. 2001. Vol. 45, No. 1. P. 5-32. DOI: 10.1023/A:1010933404324.

20. Pedregosa F., Varoquaux G., Gramfort A., Michel V., et al. Scikit-learn: machine learning in Python // Journal of Machine Learning Research. 2011. Vol. 12. P. 2825-2830.

UDC 531.55:004.94

Iliyas Arman
Grade 10 student
National School of Physics and Mathematics (RFMSh)
(Astana, Kazakhstan)

SHOOTING ON THE MOVE: HOW MUCH AIM A ROBOT LOSES BY IGNORING ITS OWN VELOCITY, AND WHETHER A VIRTUAL-TARGET CORRECTION RECOVERS IT

Abstract: A robot that launches a projectile while driving inherits its own velocity, so a firing solution computed for a stationary machine misses. Competitive robotics teams handle this with a virtual-target correction, aiming at a shifted point rather than at the goal, but the size of the error being corrected, and whether the correction survives aerodynamic drag, are rarely quantified. This study builds a point-mass trajectory model of a FIRST Tech Challenge DECODE launcher with quadratic drag and evaluates it over the competition's own launch zones, 0.86 to 3.90 metres, using a real robot's measured flywheel calibration, hood range and top speed. Three results follow. First, drag alone is not a detail: a firing solution computed without it falls short by 0.27 metres at 1.75 metres and by 0.85 metres at 3.90 metres, that is 15 to 22 per cent of the range. Second, ignoring the robot's own motion costs roughly half a metre of range per metre per second of robot speed, reaching 1.48 metres at the machine's top speed of 2.73 metres per second, which is more than twice the width of the goal opening. Third, the virtual-target correction works: aiming at a point 0.5 to 1.5 metres nearer than the goal brings the residual error to within four millimetres at every range where a solution exists. The correction has one boundary, and it is geometric rather than aerodynamic: with a turret 0.30 metres above the floor and a hood limited to 60 degrees, targets closer than about 1.5 metres cannot be entered from above at all, moving or stationary, and raising the turret to 0.60 metres removes the restriction. The practical ranking for a team is therefore drag first, own velocity second, and flywheel-speed error a distant third, since a 5 per cent speed error costs 0.22 to 0.36 metres, less than driving at one metre per second and ignoring it.

Keywords: projectile motion; aerodynamic drag; robotics; FIRST Tech Challenge; trajectory simulation; shooting on the move

Introduction

During the DECODE season, my team FoxSlide initially relied on hardcoding three specific points on the field where we could shoot while staying still. We quickly realized this restricted our gameplay and cost us valuable cycle time, so we set a new goal: to shoot accurately from anywhere on the map, even while on the move. However, firing in motion caused our shots to consistently overshoot the goal due to the robot's forward momentum. Because our brute-force approach of relying on a few inflexible, predetermined shooting spots could not keep up with the constant, unpredictable movement of teleop matches, we knew we needed a dynamic solution. I began researching projectile motion and vector addition, which led me to the concept of a virtual target. By multiplying the robot's current velocity by the ball's estimated flight time, we could calculate a shifted aim point backward on the fly, allowing the robot's momentum to carry the shot into the real goal.

A robot that throws something while driving faces a problem that a stationary one does not. The projectile leaves the launcher with the launcher's velocity added to its own, so a firing

solution computed as if the machine stood still sends the ball past the target in the direction of travel. In competitive robotics this is called shooting on the move, and the standard answer is a virtual target: instead of aiming at the goal, the robot aims at a point shifted against its direction of motion, so that the sum of the two velocities carries the projectile to the real goal.

The logic is simple, and it is usually implemented from the textbook projectile equations, in which a body follows a parabola under gravity alone. Two things about that are worth checking rather than assuming. The first is whether the parabola is good enough. A competition ball is light and large, so the ratio of aerodynamic force to weight is not negligible the way it is for a dense projectile; if drag matters, then a correction derived from the drag-free equations corrects one error while leaving another in place. The second is how large the uncorrected error actually is. Teams treat shooting on the move as an advanced feature, worth engineering effort; whether it is worth more or less than, say, tightening the tolerance on flywheel speed is an empirical question that can be answered before any code is written.

This study answers both for a concrete, fully specified machine: the FIRST Tech Challenge DECODE field, whose geometry is published in the game manual, and a real competing robot whose flywheel calibration, hood range and top speed were measured by its own team during the 2026 season. The questions are: how much range does a firing solution lose if aerodynamic drag is omitted; how much does it lose if the robot's own velocity is ignored; does the virtual-target correction remove the second error once the first is modelled; and how do both compare with the error from an imperfect flywheel.

Materials and Methods

The machine and the field. The projectile is the DECODE artifact, a moulded polypropylene ball which the game manual specifies as "5 in. (12.70 cm) nominal" and "4.9 in +/- 0.25 in. (12.45 cm +/- 0.65 cm) in diameter at the mold seam"; the mid-specification value of 0.1245 m is used. The target is the goal, whose "top lip ... is 38.75 in. (98.45 cm) from the surface of the TILE", with an opening "26.5 in. (67.30 cm) wide and 18.3 in. (46.45 cm) deep". Shots are taken from two zones which the team measured on the field: a close zone from 0.86 to 2.50 m and a far zone from 2.50 to 3.90 m.

The launcher. The robot's flywheel was calibrated in-season by its team, giving five measured points from 1050 rpm at an exit speed of 3.76 m/s to 1700 rpm at 6.04 m/s. A least-squares line through those five points is $v = 0.003679 \cdot RPM - 0.296$ (m/s), which reproduces the team's own working relation to three decimal places. The simulation uses that line over a flywheel range of 600 to 2400 rpm, that is exit speeds of 1.91 to 8.53 m/s; the portion above 1700 rpm is an extrapolation of the team's own calibration and is identified as such. The hood adjusts the launch angle between 30 and 60 degrees. The launch point is taken to be 0.30 m above the tile; the manual does not fix this, so it is an assumption, and its consequences are tested directly in the sensitivity analysis. The robot's measured top speed is 2.73 m/s.

The mass. The manual gives the artifact's size but not its mass. For a hollow polypropylene ball of this diameter with a wall of the order of 1.5 mm, the mass is about 0.049 kg, and that value is used. Because it is an estimate rather than a measurement, the whole analysis is repeated at 0.040, 0.060 and 0.080 kg, and the result is reported.

The model. The ball is treated as a point mass under gravity and quadratic aerodynamic drag, with drag coefficient 0.47 for a smooth sphere and air density 1.225 kg/m³. The equations are integrated explicitly with a 2 ms step, and a shot is scored where the trajectory crosses the plane of the goal lip while descending. A shot counts as made only if its apex clears the lip by at least 0.25 m, since a ball arriving at the rim almost horizontally grazes it rather than dropping

into a 0.46 m deep opening. Spin is not modelled: a flywheel imparts backspin and therefore lift, which would extend the range slightly, so the drag-only model is a conservative bound in that direction, and the omission is stated rather than hidden.

The firing solution. For a given target distance, the hood angle is searched over its full range in one-degree steps and the exit speed is found by bisection for each angle. Among the solutions that land on the target with the required clearance, the one closest to 45 degrees is selected, that being the standard compromise between flight time and margin. The same procedure with drag switched off gives the textbook solution, and firing the textbook solution through the drag model gives the range lost to drag.

Shooting on the move. With the stationary solution fixed, the robot's velocity is added to the horizontal component and the shot re-flown, giving the uncompensated error, at robot speeds of 0.5, 1.0, 1.5, 2.0 and 2.73 m/s. The virtual-target correction is then applied: the aim distance is adjusted iteratively until the moving shot lands on the real target, and the residual error and the required aim point are recorded.

Reproducibility. The simulation is a single Python file using only the standard library; it and its full numerical output are released with this paper, so any number below can be regenerated.

Results

Drag is not a correction, it is the dominant term at long range. A firing solution computed from the drag-free equations and then flown through the air falls short in every case, and the shortfall grows with distance: 0.271 m at a target 1.75 m away, 0.457 m at 2.50 m and 0.851 m at 3.90 m. In relative terms that is 15 per cent of the range at the short end and 22 per cent at the long end. A ball that light loses a fifth of its reach to the air, so a team that programs the parabola alone will find its far-zone shots landing consistently in front of the goal and will, in practice, calibrate the difference away by hand.

Table 1 – Stationary firing solutions with drag, and the range lost by a solution computed without it.

Target distance, m	Hood angle, °	Exit speed, m/s	Flywheel, rpm	Flight time, s	Shortfall if drag is ignored, m
1.75	57	5.34	1,532	0.66	0.271
2.25	50	5.89	1,683	0.66	0.390
2.50	47	6.21	1,767	0.66	0.457
3.00	45	6.76	1,918	0.71	0.596
3.50	45	7.27	2,055	0.79	0.730
3.90	45	7.67	2,166	0.85	0.851

Ignoring the robot's own velocity costs about half a metre per metre per second. With the stationary solution unchanged and the robot driving toward the goal, the ball overshoots by an amount almost exactly proportional to the robot's speed: at a 2.50 m target, 0.252 m at 0.5 m/s, 0.496 m at 1.0, 0.730 m at 1.5, 0.957 m at 2.0 and 1.274 m at the top speed of 2.73 m/s. The proportionality is what the physics predicts, since the extra distance is the robot's speed multiplied by the flight time, and the flight time here is about 0.66 s. The size is what matters for a team: the goal opening is 0.673 m wide, so driving at a single metre per second and not accounting for it already puts the ball outside the opening, and at top speed the miss is more than twice the opening's width.

Table 2 – Range error from ignoring the robot's own velocity (positive is an overshoot), and the residual once the virtual-target correction is applied. The aim point is where the robot must point instead of at the goal.

Target distance, m	0.5 m/s	1.0 m/s	1.5 m/s	2.0 m/s	2.73 m/s	Residual after correction, m	Aim point at 1.0 m/s, m
1.75	+0.271	+0.533	+0.786	+1.029	+1.369	no solution	—
2.25	+0.258	+0.507	+0.748	+0.979	+1.303	+0.002	1.76
2.50	+0.252	+0.496	+0.730	+0.957	+1.274	+0.003	2.01
3.00	+0.263	+0.516	+0.761	+0.997	+1.328	+0.003	2.49
3.50	+0.281	+0.552	+0.815	+1.068	+1.423	+0.001	2.96
3.90	+0.293	+0.576	+0.849	+1.113	+1.483	+0.002	3.33

The correction works, and it works to within millimetres. Applying the virtual-target correction brings the residual error to at most 3 mm at every distance and every speed at which a solution exists, including the top speed of 2.73 m/s at the far end of the field. The correction is large: at 1.0 m/s the robot must aim roughly half a metre short of the goal, and at 2.73 m/s roughly one and a half metres short. In other words, the error is worth correcting and the standard correction is sufficient; nothing more elaborate is required once drag is inside the model rather than outside it.

The correction has a boundary, and it is geometric. At the closest targets the correction has nowhere to go. At 1.75 m the compensated solution does not exist: aiming nearer than that requires a trajectory the hood cannot produce. The same restriction removes the two shortest distances, 0.90 m and 1.25 m, from the table entirely, and it is not caused by drag or by motion. With a turret 0.30 m above the floor and a maximum hood angle of 60 degrees, a ball aimed at a target 1.25 m away reaches its apex at 0.99 m, within a centimetre of the goal lip, so it arrives at the rim flat instead of dropping in. Geometrically, raising the launch point fixes it: at 0.60 m the 1.25 m target becomes reachable with the required clearance, and at 0.75 m it becomes comfortable. Whether that is available to a team is a separate question, and for this competition the answer is no. Rule R105 caps vertical expansion at 0.457 m (18 in.) in normal play, which a 0.60 m launch point already exceeds. The manual does allow expansion to 0.965 m (38 in.), but rule G415 permits it only during the final 20 seconds of a match and only outside the LAUNCH ZONES, which is precisely where short-range shots are taken. The exception therefore exists and is closed exactly where this problem lives. The near zone is a permanent blind spot to be driven around rather than a defect to be engineered out.

Table 3 – Firing solutions as a function of turret height. A dash means no solution clears the goal lip by the required margin anywhere in the 30–60° hood range.

Launch height, m	0.90 m	1.25 m	1.75 m	2.50 m	3.90 m
0.30	—	—	57°, 5.34 m/s	47°, 6.21 m/s	45°, 7.67 m/s
0.45	—	—	54°, 5.11 m/s	45°, 6.02 m/s	45°, 7.50 m/s
0.60	—	58°, 4.30 m/s	49°, 4.89 m/s	45°, 5.80 m/s	45°, 7.33 m/s
0.75	—	53°, 4.01 m/s	45°, 4.69 m/s	45°, 5.60 m/s	45°, 7.18 m/s

Sensitivities. The artifact mass is an estimate, so the central result was recomputed across 40, 49, 60 and 80 g: at a 2.50 m target with the robot at 1.5 m/s, the uncompensated error moves from 0.691 m to 0.829 m across that whole range. The ordering and the practical conclusion are

unchanged; a heavier ball is simply less affected by the air, which also slightly increases the motion error because the flight is faster. Flywheel-speed error was tested at ±5 per cent: it costs 0.223 to 0.227 m at 1.75 m, 0.305 to 0.313 m at 2.50 m and 0.358 to 0.360 m at 3.50 m. That is the comparison a team can act on. A five per cent speed error is worth about 0.3 m; driving at 1 m/s without compensating is worth about 0.5 m; driving at top speed without compensating is worth about 1.3 m.

Discussion

Three numbers summarise the result, and they come in a definite order. Aerodynamic drag costs a fifth of the range at the far end of this field and must be in the model before anything else is discussed. The robot's own velocity costs about half a metre for every metre per second of speed, which exceeds the width of the target opening at one metre per second. Flywheel-speed error, at a realistic five per cent, costs about a third of a metre. A team deciding where to spend its engineering effort should therefore model the air first, implement shooting-on-the-move second, and tighten flywheel control third, and this ordering is the practical contribution of the paper.

The virtual-target correction itself needs no defence: with drag inside the model it lands the ball within three millimetres at every distance and speed where a firing solution exists. What does deserve attention is where it stops working, because the reason is instructive. The failure at short range is not aerodynamic and not about motion; it is the geometry of a low turret firing into a raised opening. A ball launched from 0.30 m at a target 1.25 m away reaches the top of its arc at the height of the goal lip, so it skims the rim rather than entering, and no amount of correction changes that. Raising the launch point by a third of a metre would remove the restriction geometrically, but the competition's own size rules do not allow it. R105 caps vertical expansion at 0.457 m in normal play, below the height the geometry demands, and the one route above it, to 0.965 m, is gated by G415 to the last 20 seconds of a match and to positions outside the LAUNCH ZONES, which excludes the short-range case entirely. The practical conclusion is therefore the opposite of a build instruction. The short-range zone cannot be fixed by software correction or by physical extension, so it is a fixed property of the game for any robot of this architecture, and the usable response is a driving decision: treat the near zone as unshootable and position outside it. This is the kind of thing a simulation can establish before a season of trial and error spends matches discovering it.

Three limits bound these results. Backspin is not modelled, and a flywheel launcher always imparts it; the resulting lift lengthens the flight, so the true ranges are somewhat longer than the model gives and the drag shortfall in Table 1 is an upper bound. The artifact mass is estimated rather than measured, although the sensitivity analysis shows the conclusions hold across a doubling of it. And the model is two-dimensional: it treats the robot as driving directly toward the goal, whereas a real robot moves at an angle, which introduces a lateral component this design does not address. The natural next step is a wind-tunnel-free measurement of the ball's actual drag by filming a few shots and fitting the trajectory, which would replace two of these assumptions with data from the field itself.

While I expected the robot's momentum to significantly overshoot our targets, I was surprised to find that aerodynamic drag was actually our largest source of error, costing up to 22 per cent of our range. It was also unexpected to learn that our close-range misses were a strict geometric limitation caused by our 0.30 m turret height. Knowing this, I would immediately integrate a drag-compensation model into our firing code rather than focusing primarily on the robot's velocity. Most importantly, because the DECODE manual caps vertical expansion at 18

in. in normal play and only lifts that cap outside the launch zones in the last 20 seconds, raising the turret to 0.60 m is not something we are allowed to do where the short shots happen, meaning close-range shots are a permanent blind spot that our drive team must strategically avoid, since neither software corrections nor physical extensions can fix them.

Conclusion

For a FIRST Tech Challenge DECODE launcher operating over the competition's own launch zones of 0.86 to 3.90 m, a firing solution that omits aerodynamic drag falls short by 0.27 m at 1.75 m and by 0.85 m at 3.90 m, 15 to 22 per cent of the range. Ignoring the robot's own velocity produces an overshoot proportional to that velocity, 0.50 m at 1 m/s and 1.27 m at the measured top speed of 2.73 m/s at a 2.50 m target, which exceeds the 0.673 m width of the goal opening at every speed above about one metre per second. The virtual-target correction removes this error to within 3 mm at all distances and speeds where a solution exists, and it requires aiming 0.5 m short at 1 m/s and about 1.5 m short at top speed. The correction fails only at the closest targets, and for a geometric reason: from a turret 0.30 m above the floor with a hood limited to 60 degrees, targets nearer than about 1.5 m cannot be entered from above at all, while a turret at 0.60 m removes the restriction. Ranked by what they cost, the three error sources are drag, then own velocity, then flywheel-speed error at about 0.3 m for a 5 per cent deviation.

Acknowledgments

Field and projectile specifications are published by FIRST in the 2026 FTC DECODE game manual and are freely available. The flywheel calibration, hood range, launch-zone measurements and robot top speed were measured by FTC team FoxSlide #29029 during the 2026 season. The simulation code and full numerical output are released with this paper. The author declares no conflict of interest and received no funding for this work.

References:

1. FIRST. FIRST Tech Challenge DECODE Game Manual Part 2, 2026. https://ftc-resources.firstinspires.org/ftc/archive/2026/game/manual

2. Chudinov P. S. Approximate analytical description of the projectile motion with a quadratic drag force. Athens Journal of Sciences, 2014, vol. 1, no. 2, pp. 97–106. https://doi.org/10.30958/ajs.1-2-2

3. Yabushita K., Yamashita M., Tsuboi K. An analytic solution of projectile motion with the quadratic resistance law using the homotopy analysis method. Journal of Physics A: Mathematical and Theoretical, 2007, vol. 40, no. 29, pp. 8403–8416. https://doi.org/10.1088/1751-8113/40/29/015

4. Achenbach E. Experiments on the flow past spheres at very high Reynolds numbers. Journal of Fluid Mechanics, 1972, vol. 54, no. 3, pp. 565–575. https://doi.org/10.1017/S0022112072000874

5. Nathan A. M. The effect of spin on the flight of a baseball. American Journal of Physics, 2008, vol. 76, no. 2, pp. 119–124. https://doi.org/10.1119/1.2805242

6. Goodwill S. R., Chin S. B., Haake S. J. Aerodynamics of spinning and non-spinning tennis balls. Journal of Wind Engineering and Industrial Aerodynamics, 2004, vol. 92, no. 11, pp. 935–958. https://doi.org/10.1016/j.jweia.2004.05.004

УДК 004.056:378.147

Ануар Әлімжан Бахытжанұлы
магистрант
кафедра «Информационная безопасность»
Евразийский национальный университет им. Л.Н. Гумилева
(г. Астана, Казахстан)

Сексенбаева Айша Курманбековна
к.ф.-м.н., доцент
кафедра «Информационная безопасность»
Евразийский национальный университет им. Л.Н. Гумилева
(г. Астана, Казахстан)

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СТАНДАРТОВ И МЕТОДИК ОБУЧЕНИЯ КИБЕРБЕЗОПАСНОСТИ: NICE FRAMEWORK, ECSF И ПРАКТИКО-ОРИЕНТИРОВАННЫЕ ТЕХНОЛОГИИ

Аннотация: Сравниваются две ведущие рамки компетенций по кибербезопасности: американская NICE Framework (NIST SP 800-181) и европейская European Cybersecurity Skills Framework (ECSF), разработанная ENISA. Отдельно разобраны современные практико-ориентированные форматы обучения, то есть киберполигоны, изолированные учебные среды («песочницы»), CTF-соревнования, геймификация и адаптивное обучение на базе искусственного интеллекта. Актуальность подкрепляется казахстанской статистикой: за I квартал 2025 г. инцидентов информационной безопасности в стране зарегистрировано более чем вдвое больше, чем за тот же квартал 2024 г. Стандарты сопоставлялись по единому списку критериев; помимо них рассмотрены пять международных образовательных платформ и научные публикации 2019–2026 гг. NICE Framework детальнее описывает задачи, знания и навыки. С ECSF же проще проектировать образовательную программу: профилей ролей в ней всего двенадцать, и рамка увязана с e-CF и EQF. Авторы предлагают матрицу соответствия «профиль роли — ведущий метод обучения» и смешанную модель подготовки, где адаптивная теория, практика в изолированных средах и игровой контроль работают в связке. Для вузов Республики Казахстан сформулированы рекомендации по обновлению образовательных программ.

Ключевые слова: кибербезопасность, NICE Framework, ECSF, рамка компетенций, киберполигон, CTF, геймификация, адаптивное обучение, человеческий фактор, Казахстан.

Введение

Устойчивое цифровое развитие государства сегодня трудно представить без кибербезопасности. Сложность в том, что атак становится больше и они усложняются быстрее, чем успевают готовить специалистов. ОЭСР, изучив ситуацию во Франции, Германии и Польше, констатирует хроническую нехватку квалифицированных кадров в этой сфере и расхождение между тем, чему учат образовательные программы, и тем, что нужно работодателям [1].

В Казахстане проблема ощущается особенно остро. Ranking.kz, обобщив данные KZ-CERT, приводит такие цифры: в январе–марте 2025 г. зарегистрировано 29 982

инцидента информационной безопасности, а за январь–март 2024 г. их было 14 868, то есть прирост составил 101,7 % [2]. Как инциденты распределяются по типам, видно из таблицы 1.

Таблица 1 – Инциденты информационной безопасности в Республике Казахстан, январь–март, ед. (составлено авторами по [2])

Тип инцидента	2024 г.	2025 г.	Прирост, %	Доля в 2025 г., %
Ботнеты	1 748	17 606	+907,2	58,7
Вирусы, сетевые черви, трояны	9 586	7 870	−17,9	26,2
Фишинг	1 439	1 975	+37,2	6,6
Отсутствие доступа к интернет-ресурсу	216	112	−48,1	0,4
DoS/DDoS-атаки	30	23	−23,3	0,1
Несанкционированный доступ и модификация содержания	13	9	−30,8	0,0
Прочие инциденты	1 836	2 387	+30,0	8,0
Всего	14 868	29 982	+101,7	100,0

Сильнее всего выросли ботнет-активность и фишинг, а успех этих угроз во многом зависит от того, как действуют пользователи и администраторы. Среди стран СНГ на Казахстан приходится около 8 % успешных атак; доля DDoS-атак в целом по СНГ составляет 18 %, что более чем вдвое выше среднемировых 8 % [3]. Авторы глобальных обзоров отмечают, что злоумышленники всё чаще автоматизируют атаки и подключают инструменты искусственного интеллекта, из-за чего учебное содержание приходится постоянно обновлять [4; 5].

Самым уязвимым звеном защиты по-прежнему остаётся человеческий фактор. Социальная инженерия эксплуатирует доверие, любопытство и готовность помочь [6]. Как выяснили Pahnila и соавт., соблюдение сотрудниками политик ИБ определяется не одними знаниями: свою роль играют установки, привычки и намерение следовать правилам [7]. Значит, обучающимся нужно прививать устойчивые модели поведения, а для этого требуются практико-ориентированные методы. На национальном уровне задачи развития отрасли закреплены в Концепции цифровой трансформации, развития отрасли ИКТ и кибербезопасности на 2023–2029 гг. [8].

Разрабатывая образовательные программы, вузы ориентируются на международные стандарты компетенций, в первую очередь на NICE (США) и ECSF (ЕС). По наблюдениям авторов, учебная база при этом отстаёт от того, как меняются киберугрозы: лабораторные практикумы и инструменты быстро устаревают. Работ, в которых эти рамки сопоставлялись бы между собой и увязывались с актуальными педагогическими методиками, в отечественной литературе пока немного.

Цель исследования заключается в обосновании модели практико-ориентированной подготовки специалистов по кибербезопасности для вузов Республики Казахстан. Модель строится на сравнении рамок компетенций NICE и ECSF и современных методов обучения.

Задачи исследования:

1) сопоставить NICE Framework и ECSF по общему набору критериев;
2) изучить ведущие международные образовательные платформы;
3) обобщить данные об эффективности практико-ориентированных методов обучения;
4) соотнести профили ролей с методами обучения и предложить гибридную модель подготовки.

Материалы и методы

Работа имеет аналитический характер и опирается на три метода, которые дополняют друг друга.

1. Сравнительный документальный анализ стандартов. В качестве объектов анализа взяты публикация NIST SP 800-181 Rev. 1 [9], действующие компоненты NICE Framework [10], а также профили ролей и руководство пользователя ECSF [11; 12]. Документы сравнивались по следующим критериям: разработчик и статус документа, его структура и единица описания, количество ролей, степень детализации, связь с квалификационными системами, наличие инструментов для разработки учебных программ и ограничения, которые возникают при применении в Казахстане.

2. Сравнительный анализ образовательных платформ. Пять международных платформ (Cybrary, Coursera, Udemy, Hack The Box, TryHackMe) сравнивались по основной специализации, формату контента, модели доступа и наличию практической среды. Информация взята из открытых описаний самих платформ.

3. Обзор научных публикаций. В выборку вошли работы 2019–2026 гг. из Scopus, IEEE Xplore, SpringerLink и ScienceDirect, найденные по запросам «cybersecurity education», «gamification», «cyber range», «security awareness», а также классические труды по теории обучения. Публикации без эмпирических данных или теоретического обоснования методов в анализ не включались.

По итогам трёх этапов составлена сводная матрица «профиль роли — метод обучения», и уже на её основе предложена гибридная модель подготовки.

Теоретические основы и обзор литературы

Теоретическую основу практико-ориентированного обучения задаёт модель экспериментального обучения Д. Колба. По Колбу, прочное знание складывается в цикле «конкретный опыт — рефлексия — концептуализация — активное экспериментирование» [13]. Из теории когнитивной нагрузки Дж. Свеллера следует, что сложность заданий нужно дозировать, а на первых этапах полезно опираться на пошаговые примеры [14]. Андрагогический подход М. Ноулза исходит из того, что взрослые лучше усваивают материал, который напрямую связан с их профессиональными задачами [15]. Мотивацию объясняет теория самодетерминации Э. Деси и Р. Райана: внутренняя мотивация сохраняется, когда удовлетворены потребности в компетентности, автономии и причастности [16].

По данным систематического обзора Hamari и соавт., геймификация в целом даёт положительный эффект, однако он зависит от контекста и от особенностей пользователей [17]. Для сферы информационной безопасности Pahlavanpour и Gao в систематическом картировании показали, что исследований геймифицированных программ повышения осведомлённости становится больше, но строгих оценок их долгосрочного эффекта по-прежнему не хватает [18].

Эмпирические работы эту неоднозначность подтверждают. В эксперименте Weijsenfeld и Sarmah у студентов, изучавших стеганографию через сюжетную игру,

вовлечённость была значимо выше, однако по результатам теста знаний статистически значимой разницы с традиционным форматом не обнаружилось [19]. Mpanza, Gundu и Futcher сформулировали практические рекомендации по применению геймификации против социальной инженерии [20]. Veneruso и соавт. показали, как виртуальную реальность можно использовать для обучения безопасному поведению [21]. По данным Fertig и соавт., e-learning-платформы помогают повысить осведомлённость в вопросах информационной безопасности [22], а Hijji и Alam разработали рамку обучения для сотрудников, работающих удалённо [23].

О разрыве между академической подготовкой и запросами практики говорят давно: ещё в 2000-х гг. исследователи предлагали сместить акцент программ на практические навыки и разрабатывать учебные планы вместе с отраслью [24; 25].

Из обзора следуют три вывода. Во-первых, практико-ориентированные и игровые методы стабильно повышают вовлечённость. Во-вторых, их влияние на объём знаний неоднозначно и определяется тем, как выстроено обучение. В-третьих, ни один метод сам по себе не охватывает всех компетенций, поэтому методы нужно обоснованно сочетать и привязывать к профессиональным ролям.

Результаты

1. Сравнение рамок компетенций NICE и ECSF

Итоги сравнения представлены в таблице 2.

Таблица 2 – Сравнение NICE Framework и ECSF (составлено авторами по [9–12])

Критерий	NICE Framework (США)	ECSF (Европейский союз)
Разработчик, год	NIST; SP 800-181 (2017), Rev. 1 (2020); компоненты обновляются регулярно (v1.0.0 — 2024 г., v2.2.0 — апрель 2026 г.)	ENISA; сентябрь 2022 г.
Структура	7 категорий рабочих ролей → рабочие роли → утверждения о задачах, знаниях и навыках (TKS); области компетенций	12 профилей ролей по единому шаблону
Единица описания	Отдельное утверждение TKS (модульный «конструктор»)	Профиль роли: миссия, результаты, задачи, ключевые навыки и знания, е-компетенции
Объём (v1.0.0)	52 рабочие роли, 11 областей компетенций, около 2 280 утверждений TKS	12 профилей
Связь с квалификациями	Прямой связи с уровнями квалификаций нет	Связь с e-CF (EN 16234-1, уровни e-1…e-5) и через неё с EQF
Инструменты для вузов	Программа центров компетенций (CAE), каталоги курсов NICCS	Проекты SPARTA (конструктор учебных программ) и CONCORDIA (каталог курсов)

Критерий	NICE Framework (США)	ECSF (Европейский союз)
Сильные стороны	Высокая детализация, точная привязка заданий к задачам	Компактность, понятность для работодателей и вузов, совместимость с европейскими квалификациями
Ограничения для РК	Избыточна для бакалавриата; часть ролей связана со спецификой госсектора США	Меньшая детализация; нет готовых измеримых индикаторов для оценки

Получается, что рамки дополняют друг друга. ECSF удобно применять на уровне образовательной программы, чтобы задать траектории выпускника, а NICE лучше подходит для уровня дисциплин и отдельных заданий, где проверяемые навыки нужно сформулировать точно.

2. Сравнение образовательных платформ

Таблица 3 – Сравнение международных платформ обучения кибербезопасности

Платформа	Специализация	Формат контента	Модель доступа	Практическая среда
Cybrary	Подготовка к сертификации	Видеолекции, лабораторные	Подписка (есть бесплатная часть)	Виртуальные лаборатории
Coursera	Академические курсы вузов	Курсы, специализации	Оплата курса или подписка	Проекты в браузере
Udemy	Прикладные навыки от практиков	Видеоуроки, тесты	Разовая покупка курса	Зависит от автора
Hack The Box	Наступательная безопасность, CTF	Уязвимые машины, киберполигоны	Freemium / подписка	Реальные цели через VPN
TryHackMe	Сценарное обучение (learning paths)	Интерактивные «комнаты»	Freemium / подписка	Встроенные виртуальные машины

Платформы с игровыми механиками (Hack The Box, TryHackMe) сразу дают обратную связь и добавляют элемент соревнования, но управленческие и правовые компетенции почти не затрагивают. Академические платформы (Coursera, Udemy) хорошо закладывают теорию, а практики в них немного. Весь спектр ролей ECSF не охватывает ни одна из платформ.

3. Матрица методов обучения

Таблица 4 – Сравнительная характеристика практико-ориентированных методов

Метод	Ключевой механизм	Основное преимущество	Ограничение
Геймификация	Баллы, уровни, рейтинги, сюжет	Рост вовлечённости и мотивации [17–19]	Эффект на знания неоднозначен; риск упрощения [19]
Песочница (sandbox)	Изолированная среда выполнения	Безопасный анализ вредоносного ПО и ошибок	Требования к инфраструктуре
Киберполигон (cyber range)	Моделирование атак на инфраструктуру	Навыки реагирования и командной работы	Сложность и стоимость сценариев
CTF-соревнования	Задачи с «флагом», ограничение по времени	Объективная оценка навыков в стрессовых условиях	Смещение в сторону технических ролей
Адаптивное обучение (ИИ)	Подстройка сложности под обучающегося	Индивидуальный темп, контроль когнитивной нагрузки [14]	Нужны данные и качественный контент

Сопоставление профилей ECSF (вместе с соответствующими категориями NICE) и методов обучения позволило построить матрицу соответствия (таблица 5). Она и составляет основной авторский результат исследования.

Таблица 5 – Матрица соответствия «профиль роли — ведущий метод обучения» (разработано авторами)

Профиль ECSF	Категория NICE	Ведущий метод	Форма контроля
Cyber Incident Responder	Protection and Defense	Киберполигон (учения Red/Blue Team)	Командные учения по сценарию
Penetration Tester	Protection and Defense	CTF, платформы типа Hack The Box	CTF, отчёт о тестировании
Digital Forensics Investigator	Investigation	Песочница, анализ образов	Кейс с заключением эксперта
Threat Intelligence Specialist	Cyberspace Intelligence	Кейс-метод, работа с источниками об угрозах	Аналитический отчёт
Cybersecurity Implementer, Architect	Implementation and Operation; Design and Development	Лабораторные в песочнице, проектная работа	Защита проекта
Risk Manager, Auditor, Legal/Compliance Officer	Oversight and Governance	Деловые игры, кейс-метод	Аудит учебной организации
CISO	Oversight and Governance	Штабные учения (tabletop)	Разбор управленческих решений

Обсуждение

Результаты согласуются с выводами обзора [18]: игровые и практические методы прежде всего вовлекают обучающихся, а прирост знаний появляется тогда, когда они сочетаются с систематической теорией и рефлексией. Исходя из этого, предлагается **гибридная модель подготовки**, состоящая из трёх блоков:

– **теоретический блок** строится на адаптивных модулях LMS: они выравнивают уровень обучающихся и позволяют управлять когнитивной нагрузкой [14];
– **практический блок** проходит в песочницах и на киберполигонах, где обучающиеся отрабатывают сценарии под свой целевой профиль (таблица 5);
– **контрольный блок** проводится в формате CTF-соревнований и командных учений, что даёт объективную оценку готовности действовать в условиях реальной атаки.

Практический блок организован по циклу Колба [13] (рисунок 1). Ошибка в изолированной среде ничем не грозит и сама становится материалом для обучения.



Рисунок 1 – Цикл практико-ориентированного обучения в изолированной среде (разработано авторами на основе [13])

Мотивационно-адаптивный контур модели (рисунок 2) связывает геймификацию с аналитикой успеваемости. Результаты выполненных заданий поступают в адаптивный модуль, и он подбирает следующее задание под уровень конкретного обучающегося. Так поддерживается и потребность в компетентности [16], и оптимальная сложность заданий [14].

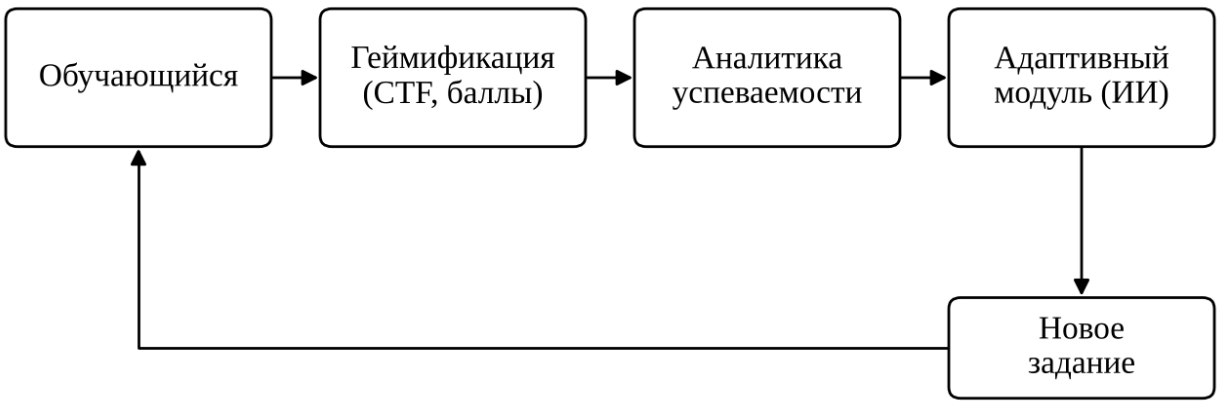


Рисунок 2 – Мотивационно-адаптивный контур гибридной модели (разработано авторами)

Ограничения исследования. Исследование носит аналитический характер, и эффективность предложенной модели экспериментально не проверялась. Платформы сравнивались по их открытым описаниям, а соотнесение ролей и методов отражает экспертное мнение авторов. Дальнейшая работа предполагает апробацию модели в учебном процессе на контрольной и экспериментальной группах.

Практические рекомендации

Вузам Казахстана, которые ведут программы по информационной безопасности, можно рекомендовать следующее:

1) описывать результаты обучения по образовательной программе через профили ECSF, а задания в отдельных дисциплинах формулировать на основе утверждений TKS из NICE;
2) сделать работу в изолированных средах (облачных песочницах, на киберполигонах) обязательной частью технических дисциплин;
3) часть привычных тестов заменить CTF-заданиями и командными учениями, оставив письменный разбор решений для развития рефлексии;
4) встраивать адаптивные модули в университетские LMS, чтобы выстраивать индивидуальные траектории обучения;
5) в подготовке по управленческим и правовым профилям дополнять технические платформы деловыми играми и штабными учениями;
6) опираться на национальную статистику инцидентов (ботнеты, фишинг) при разработке учебных сценариев.

Заключение

Анализ показал, что NICE Framework и ECSF друг друга дополняют. ECSF удобна, когда нужно спроектировать образовательную программу и согласовать её с рынком труда, а NICE подходит для детальной проработки заданий и оценочных средств. Обзор литературы подтвердил, что практико-ориентированные и игровые методы повышают вовлечённость, но их влияние на знания зависит от того, сочетаются ли они с теорией и рефлексией.

Научная новизна исследования связана с двумя результатами: матрицей соответствия профилей ролей ECSF и категорий NICE ведущим методам обучения и гибридной моделью подготовки, которая объединяет адаптивный теоретический блок, практику в изолированных средах и геймифицированный контроль. Практическая значимость состоит в том, что результаты можно использовать при обновлении образовательных программ по информационной безопасности в казахстанских вузах. В перспективе планируется проверить модель экспериментально и разработать показатели для оценки её эффективности.

Список литературы:

1. Building a Skilled Cyber Security Workforce in Europe: Insights from France, Germany and Poland / D. Eslava [et al.]. – Paris : OECD Publishing, 2024. – 114 p. – (OECD Skills Studies). – DOI: 10.1787/3673cd60-en.

2. Количество кибератак в Казахстане выросло в два раза [Электронный ресурс] // Капитал.kz. – 2025. – 2 мая. – URL: https://kapital.kz/tehnology/136625/kolichestvo-kiberatak-v-kazahstane-vyroslo-v-dva-raza.html (дата обращения: 06.10.2026).

3. Россия под прицелом: 75 % кибератак в СНГ нацелены на одну страну [Электронный ресурс] // SecurityLab.ru. – 2024. – URL: https://www.securitylab.ru/news/551770.php (дата обращения: 06.10.2026).

4. Gajjar V. R., Taherdoost H. Cybercrime on a Global Scale: Trends, Policies, and Cybersecurity Strategies // 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI). – IEEE, 2024. – P. 668–676.

5. Influence of Digital Technologies on Information Security in the Public Administration System / M. Durman [et al.] // Journal of Theoretical and Applied Information Technology. – 2025. – Vol. 103, No. 2. – P. 400–411.

6. Human Factor, a Critical Weak Point in the Information Security of an Organization's Internet of Things / K. Hughes-Lartey [et al.] // Heliyon. – 2021. – Vol. 7, No. 3. – Art. e06522. – DOI: 10.1016/j.heliyon.2021.e06522.

7. Pahnila S., Siponen M., Mahmood A. Employees' Behavior towards IS Security Policy Compliance // Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS'07). – IEEE, 2007. – P. 156b.

8. Zhamburbayeva S., Ilyassova G. A. The Realization of the "Concept of Digital Transformation, Development of the Information and Communication Technologies and Cybersecurity Industry for 2023–2029" by Implementing Blockchain in the Technologies of the Republic of Kazakhstan and the Problems of Its Legal Regulation // Вестник Карагандинского университета. Серия «Право». – 2024. – Т. 29, № 4. – С. 137–145.

9. Workforce Framework for Cybersecurity (NICE Framework) : NIST Special Publication 800-181 Rev. 1 / R. Petersen [et al.]. – Gaithersburg : NIST, 2020. – DOI: 10.6028/NIST.SP.800-181r1.

10. NICE Framework Latest Updates [Электронный ресурс] // National Institute of Standards and Technology. – URL: https://www.nist.gov/itl/applied-cybersecurity/nice/nice-framework-resource-center/nice-framework-latest-updates (дата обращения: 06.10.2026).

11. European Cybersecurity Skills Framework (ECSF): Role Profiles [Электронный ресурс] / European Union Agency for Cybersecurity (ENISA). – 2022. – URL: https://www.enisa.europa.eu/publications/european-cybersecurity-skills-framework-role-profiles (дата обращения: 06.10.2026).

12. European Cybersecurity Skills Framework (ECSF): User Manual [Электронный ресурс] / European Union Agency for Cybersecurity (ENISA). – 2022. – URL: https://www.enisa.europa.eu/topics/cybersecurity-education/european-cybersecurity-skills-framework (дата обращения: 06.10.2026).

13. Kolb D. A. Experiential Learning: Experience as the Source of Learning and Development. – 2nd ed. – Upper Saddle River : Pearson FT Press, 2014. – 416 p.

14. Sweller J. Cognitive Load Theory // Psychology of Learning and Motivation. – 2011. – Vol. 55. – P. 37–76. – DOI: 10.1016/B978-0-12-387691-1.00002-8.

15. Knowles M. S., Holton E. F., Swanson R. A. The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development. – 8th ed. – London : Routledge, 2014. – DOI: 10.4324/9781315816951.

16. Deci E. L., Ryan R. M. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior // Psychological Inquiry. – 2000. – Vol. 11, No. 4. – P. 227–268. – DOI: 10.1207/S15327965PLI1104_01.

17. Hamari J., Koivisto J., Sarsa H. Does Gamification Work? A Literature Review of Empirical Studies on Gamification // Proceedings of the 47th Hawaii International Conference on System Sciences. – IEEE, 2014. – P. 3025–3034. – DOI: 10.1109/HICSS.2014.377.

18. Pahlavanpour O., Gao S. A Systematic Mapping Study on Gamification within Information Security Awareness Programs // Heliyon. – 2024. – Vol. 10, No. 19. – Art. e38474. – DOI: 10.1016/j.heliyon.2024.e38474.

19. Weijsenfeld F. G. J., Sarmah D. K. Gamifying Cybersecurity: A Narrative-Driven Approach to Teaching Steganography // Computers and Education Open. – 2025. – Vol. 9. – Art. 100288. – DOI: 10.1016/j.caeo.2025.100288.

20. Mpanza A., Gundu T., Futcher L. Gamified Defence: Practical Guidelines for Combating Social Engineering Attacks // Information Security Education. Empowering People Through Information Security Education : WISE 2025. – Cham : Springer, 2026. – (IFIP Advances in Information and Communication Technology ; vol. 742).

21. CyberVR: An Interactive Learning Experience in Virtual Reality for Cybersecurity-Related Issues / S. Veneruso [et al.] // Proceedings of the International Conference on Advanced Visual Interfaces (AVI ’20). – New York : ACM, 2020.

22. Measuring the Impact of E-Learning Platforms on Information Security Awareness / T. Fertig [et al.] // Learning and Collaboration Technologies : HCII 2019. – Cham : Springer, 2019. – (Lecture Notes in Computer Science ; vol. 11591). – P. 26–37. – DOI: 10.1007/978-3-030-21814-0_3.

23. Hijji M., Alam G. Cybersecurity Awareness and Training (CAT) Framework for Remote Working Employees // Sensors. – 2022. – Vol. 22, No. 22. – Art. 8663. – DOI: 10.3390/s22228663.

24. Hentea M., Dhillon H. S., Dhillon M. Towards Changes in Information Security Education // Journal of Information Technology Education: Research. – 2006. – Vol. 5. – P. 221–233.

25. Armstrong H., Jayaratna N. Internet Security Management: A Joint Postgraduate Curriculum Design // Journal of Information Systems Education. – 2002. – Vol. 13, No. 3. – P. 249–258.

UDC 004.41:330.341.1

Mukhamedzhan Yerganat
First-year student
Data Analytics and AI Management
Kazakh-British Technical University
(Almaty, Kazakhstan)
Scientific supervisor: Askarbekkyzy Aknur
Lecturer, Kazakh-British Technical University
(Almaty, Kazakhstan)

GROWTH OF GITHUB DEVELOPER COMMUNITIES IN KAZAKHSTAN, UZBEKISTAN AND KYRGYZSTAN, 2020-2026: EVIDENCE FROM THE GITHUB INNOVATION GRAPH

Abstract: This study asks how the number of software developers on GitHub changed in Kazakhstan and Uzbekistan from 2020 to 2026, which country grew faster, and whether programming languages such as Python grew at the same pace as the developer base. Kyrgyzstan is added as a third Central Asian country. The data are the open GitHub Innovation Graph, which reports quarterly counts of developer accounts, repositories, git pushes and language users for each economy, for 25 quarters from 2020 Q1 to 2026 Q1, combined with World Bank population figures. Developer accounts in Kazakhstan rose from 69,649 to 463,403 (6.7 times), in Uzbekistan from 24,575 to 395,040 (16.1 times, the fastest growth among the 121 economies that had at least 10,000 accounts in 2020 Q1) and in Kyrgyzstan from 14,440 to 121,507 (8.4 times). Per million residents Kazakhstan still leads, with 22,232 accounts against 10,661 in Uzbekistan, but its lead fell from 4.9 times to 2.1 times. Yearly git pushes grew in step with accounts in Kazakhstan and Uzbekistan but not in Kyrgyzstan. In Kazakhstan and Uzbekistan the number of developers pushing to Python repositories grew faster than both the developer base and the number of pushes. Kazakhstan's code activity drops in every third quarter: in 2020-2025 git pushes were on average 19.6% below the level of the neighbouring quarters (about 15% below trend) and Python users 28.5% below, while Uzbekistan shows a much smaller drop (3.5% in git pushes, not significant, and 8.7% in Python users). An exploratory comparison of 146 economies shows that this summer drop is common in the Northern Hemisphere and weaker or reversed in the Southern Hemisphere, and the north-south contrast is sharper outside the tropics, so it is consistent with the northern summer holiday season rather than a problem specific to Kazakhstan's IT sector.

Keywords: GitHub, software developers, open data, Kazakhstan, Uzbekistan, Kyrgyzstan, Python, seasonality, digital economy.

Introduction

Software is now a visible part of the economies of Central Asia. Kazakhstan reported 1.142 billion US dollars of IT service exports in 2025 [11], and the official statistics of Uzbekistan show that the service exports of IT Park participants grew from 276.3 million US dollars in 2022 to 728.4 million in 2025 [10]. Such figures say little about the people who write code: how many there are, how fast their number grows and which technologies they use.

Data from GitHub, the most widely used public platform for git projects [3], give a direct trace of these people. Researchers have used GitHub data to map where open source developers live [3], to measure the software capabilities of countries [4] and to estimate the effect of new

tools on programming activity [5]. GitHub also publishes the Innovation Graph, an open dataset that counts developer accounts, repositories, code uploads (git pushes) and users of each programming language for every economy and every quarter since 2020 [1, 2].

A first look at this dataset shows a clear picture for Central Asia. At the start of 2020 Kazakhstan had 69,649 developer accounts on GitHub and Uzbekistan had 24,575, about a third as many. By the first quarter of 2026 Kazakhstan had 463,403 and Uzbekistan 395,040, so Uzbekistan had almost caught up. This paper asks a simple question about that picture: how did the number of developers in Kazakhstan and Uzbekistan change on GitHub since 2020, which country grows faster, and do different programming languages, for example Python, grow at the same pace? Kyrgyzstan is added as a third country, as a smaller neighbour with a similar post-Soviet background.

Two earlier papers in this journal compared Kazakhstan with Uzbekistan using official statistics on inflation, unemployment and taxes [7, 8]. This paper uses the same side-by-side design, but applies it to open data on software development. Among the studies reviewed for this paper, Kazakhstan and Uzbekistan appear only as rows in cross-country tables [4, 6]; none of them follows the two countries quarter by quarter.

Literature review

Wachs, Nitecki, Schueller and Polleres located more than half a million active GitHub contributors in early 2021, mostly from the location that users write in their profiles, and compared the result with earlier snapshots [3]. The share of developers in Asia, Latin America and Eastern Europe grew, while the share of the United States fell from 38.7% in 2010 to 24.6% in 2021.

Juhász, Wachs, Kaminski and Hidalgo used the Innovation Graph, in which a developer is assigned to a country by IP address rather than by a self-reported location, for 163 countries and 150 programming languages in 2020-2023 [4]. From the mix of languages each country uses they built a software economic complexity index and showed that it helps to explain differences in income per person, inequality and emissions. In their 2020 ranking Uzbekistan is 70th, Kazakhstan 71st and Kyrgyzstan 92nd among the 126 countries listed (supplementary Tables S6.1-S6.2 of the arXiv version of [4]). By this measure the two neighbours started the period at almost the same level.

Quispe and Grijalba used the same dataset to compare countries where ChatGPT was available with countries where it was not [5]. With difference-in-differences and synthetic control methods they found a positive effect on git pushes, repositories and developers per 100,000 people, strongest for high-level general-purpose languages such as Python and JavaScript and for shell scripting.

Two more sources help to read the time pattern of the data. Wachs studied developers who left Russia after the invasion of Ukraine in 2022 [6]. His sample consists of highly active developers, who made at least 100 commits to public projects in 2019-2020. By November 2022, 11.1% of those previously located in Russia listed a new country, against 2.8% in comparable countries. Arrivals from Russia and Belarus equalled 13% of the highly active developers he had located in Kazakhstan in 2021, 8% in Uzbekistan and 7% in Kyrgyzstan (Table 3 of [6]). Finally, GitHub's own analysis of the timing of contributions reported changes in activity "during different holiday seasons and (sometimes correlatingly) when students are out of school" [12].

None of these studies follows Kazakhstan, Uzbekistan and Kyrgyzstan over time or checks whether their growth in accounts is matched by growth in actual code.

Data and methods

Data. The main source is the GitHub Innovation Graph, version 1.0.11 (repository state of 1 September 2026, downloaded on 11 October 2026) [1]. It is released under the CC0 licence, which allows any use. Four files were used. The developers file counts developer accounts located in an economy on the last day of each quarter; it excludes bots and spam accounts but includes accounts that are no longer active. The repositories file counts public software projects at the end of each quarter. The git pushes file counts how many times developers in an economy uploaded code during the quarter. The languages file counts, for each programming language, the unique developers who made at least one push to a repository in that language during the quarter [2]; below, "Python users" means developers who pushed to at least one repository in which GitHub detects Python. GitHub assigns a developer to an economy by the most frequent location of his or her IP address in the quarter and does not publish a number when fewer than 100 developers performed the activity [2]. The data cover 25 quarters, from 2020 Q1 to 2026 Q1. The author also viewed and saved the git push charts for Kazakhstan and Uzbekistan on the Innovation Graph website.

Population comes from the World Bank indicator SP.POP.TOTL (mid-year population) for 2020-2025 [9]. Each quarter uses the population of its calendar year; the quarter 2026 Q1 uses the 2025 value, the latest published. In 2025 the populations were 20,843,754 in Kazakhstan, 37,053,428 in Uzbekistan and 7,343,064 in Kyrgyzstan.

Plan. The analysis plan was written and agreed on 6 October 2026, after the first two charts (Figures 1 and 4 for Kazakhstan and Uzbekistan) had been drawn, but before the per-resident comparison and the check were computed. It had two parts. First, developer accounts per quarter were divided by population to compare the countries per million residents, to see who grows faster and in which years the gap between Kazakhstan and Uzbekistan widened. Second, the same comparison was repeated for repositories and git pushes, to check that the growth is not only new accounts without code. The language comparison follows the second half of the research question.

Measures. Growth is reported as a multiple (last value divided by first) and as a compound annual growth rate over the six years from 2020 Q1 to 2026 Q1; git pushes and language users, which change with the season, are compared as yearly totals or averages, 2020 against 2025. Activity per account is the number of git pushes in a year divided by the average number of developer accounts in that year. Comparisons with "all economies" use every economy in the Innovation Graph except the European Union aggregate, so they include Taiwan, which is not in the World Bank country list.

Exploratory analyses. Three analyses were added after a third-quarter dip was noticed in the charts, so they are exploratory and not part of the original plan; their p-values are descriptive. (1) Third-quarter test: for every year the value of the third quarter was compared with the midpoint of the second and fourth quarters on a logarithmic scale, $d = \ln(Q3) - [\ln(Q2) + \ln(Q4)] / 2$. A smooth series gives d close to zero, a negative d is a dip, and $(e^{d} - 1) \times 100$ gives its size in percent. The mean of d over the six years 2020-2025 was tested against zero with a two-sided one-sample t-test; for the nine tests in Table 4 Holm-adjusted p-values are also given. Because a peak in the second quarter also lowers d, the second and third quarters were in addition compared with a centred 2×4 moving average of the logarithms (second quarters 2021-2025, third quarters 2020-2025). (2) The same measure was computed for git pushes in every economy with all 25 quarters reported and a capital city with a latitude in the World Bank list: 146 economies (Taiwan, Réunion and the West Bank and Gaza have full series but no latitude

there and are excluded), and the same measure was computed for the first quarter (2021-2025, five years, since 2020 Q1 has no previous quarter). Northern and Southern Hemisphere economies were compared with the Mann-Whitney U test, also for economies outside the tropics only (capital latitude of at least 23.44 degrees north or south), and the size of the dip was correlated with growth in pushes (Spearman correlation). (3) The growth of accounts in 2022 Q4 in each country was compared with the same country's growth in the other quarters of 2021-2023 and with the growth of all economies together; the change in git pushes in 2022 Q4 was compared with the same country's fourth quarters of 2020-2025 and with all economies; and the jump of 2026 Q1 was compared with all economies that had at least 20,000 pushes in 2025 Q4.

Software. The calculations were made in Python 3.13 with pandas 3.0, SciPy 1.18 and matplotlib 3.11.

Results

Developer accounts: Uzbekistan grows faster, Kazakhstan stays ahead

In Figure 1 the developer accounts of all three countries rise in every quarter, but at different speeds. From 2020 Q1 to 2026 Q1 accounts in Kazakhstan grew 6.65 times, in Kyrgyzstan 8.41 times and in Uzbekistan 16.07 times, which is 37.1%, 42.6% and 58.9% per year (Table 1). Among the 121 economies that had at least 10,000 developer accounts in 2020 Q1, Uzbekistan's growth was the fastest; Kyrgyzstan was 8th, Kazakhstan 27th, and the median economy grew 5.06 times.

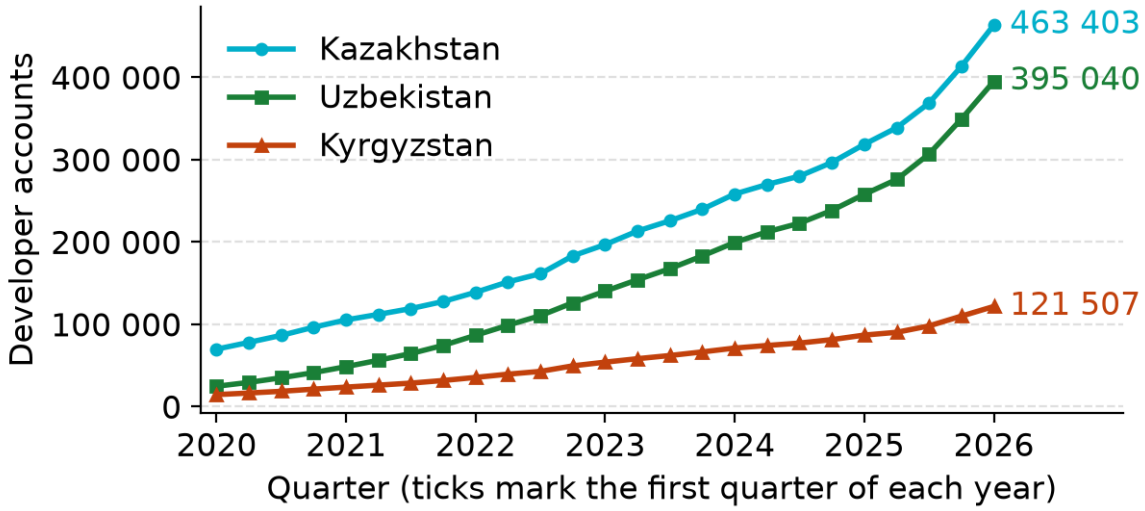


Figure 1 - Developer accounts on GitHub by quarter, 2020 Q1 to 2026 Q1 (source: GitHub Innovation Graph [1]; first version for Kazakhstan and Uzbekistan produced by the author, Kyrgyzstan added)

Uzbekistan grew faster than Kazakhstan in every single year, but the difference became smaller. On average, Uzbekistan's accounts grew by 18.7% per quarter in 2020 against 11.4% in Kazakhstan, and by 10.0% per quarter against 8.7% in 2025.

Table 1 - Size and growth of GitHub activity in Kazakhstan (KZ), Uzbekistan (UZ) and Kyrgyzstan (KG)

Indicator	KZ	UZ	KG
Developer accounts, 2020 Q1	69,649	24,575	14,440
Developer accounts, 2026 Q1	463,403	395,040	121,507
Growth, times (per year)	6.65 (37.1%)	16.07 (58.9%)	8.41 (42.6%)

Accounts per million residents, 2026 Q1	22,232	10,661	16,547
Repositories, 2020 Q1 → 2026 Q1	132,186 → 944,642	49,782 → 1,254,862	44,258 → 415,407
Git pushes per year, 2020 → 2025	433,047 → 1,813,424	190,401 → 1,837,684	174,311 → 490,975
Growth of yearly pushes, times	4.19	9.65	2.82
Growth of accounts, 2020 Q4 → 2025 Q4, times	4.29	8.49	5.21
Pushes per account per year, 2020 → 2025	5.24 → 5.04	5.86 → 6.18	9.92 → 5.11
Python users per quarter (yearly mean), 2020 → 2025	1,452 → 7,419	453 → 6,407	467 → 1,922

Note. Source: GitHub Innovation Graph [1], World Bank [9]; calculations for this study. Pushes and Python users are compared as yearly totals or yearly means because they change with the season.

Per million residents: the gap widened, the ratio narrowed

Uzbekistan has almost twice as many people as Kazakhstan, so the raw numbers hide a large difference in density (Figure 2 and Table 2). In 2020 Q1 Kazakhstan had 3,575 developer accounts per million residents and Uzbekistan 732, a ratio of 4.89. By 2026 Q1 the figures were 22,232 and 10,661, and the ratio had fallen to 2.09. Kyrgyzstan stood between the two, with 16,547 accounts per million.

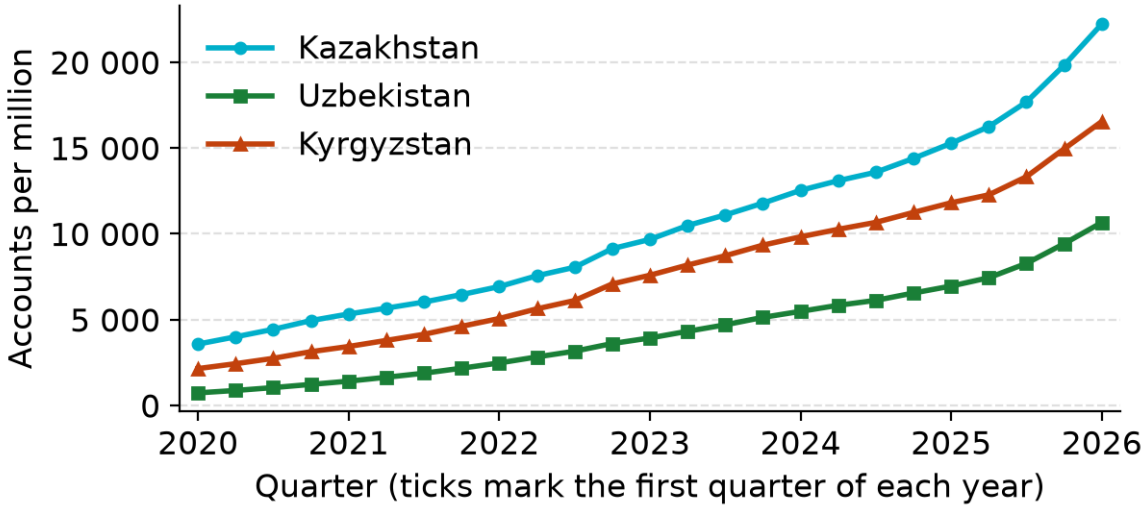


Figure 2 - Developer accounts per million residents (source: GitHub Innovation Graph [1], World Bank [9]; calculations for this study)

The answer to the question "in which year did the gap begin to grow" depends on how the gap is measured. The absolute gap, Kazakhstan minus Uzbekistan per million residents, grew in every year, from 3,719 at the end of 2020 to 10,399 at the end of 2025. It grew most in 2022 (by 1,238) and in 2025 (by 2,554). The relative gap moved the other way: Kazakhstan's lead shrank every year, from 4.04 times at the end of 2020 to 2.10 times at the end of 2025. In other words, Uzbekistan is catching up in proportion, while Kazakhstan, starting from a higher base, still adds more accounts per resident each year.

Table 2 - Developer accounts per million residents at the end of each year

Quarter	KZ	UZ	KG	Gap KZ − UZ	Change of the gap	Ratio KZ / UZ
2020 Q4	4,944	1,225	3,132	3,719	n/a	4.04
2021 Q4	6,460	2,168	4,610	4,291	+573	2.98
2022 Q4	9,135	3,606	7,079	5,529	+1,238	2.53
2023 Q4	11,787	5,131	9,339	6,657	+1,127	2.30
2024 Q4	14,396	6,550	11,247	7,845	+1,189	2.20
2025 Q4	19,825	9,425	14,953	10,399	+2,554	2.10
2026 Q1	22,232	10,661	16,547	11,571	+1,171	2.09

Note. Source: GitHub Innovation Graph [1], World Bank [9]; calculations for this study. The change of the gap is computed from unrounded values; for 2026 Q1 it is the change over one quarter.

The check: is there code behind the accounts?

The developer count includes unused accounts, so it could grow even if nobody writes more code. The planned check therefore looked at repositories and git pushes (Figure 3 and Table 1).

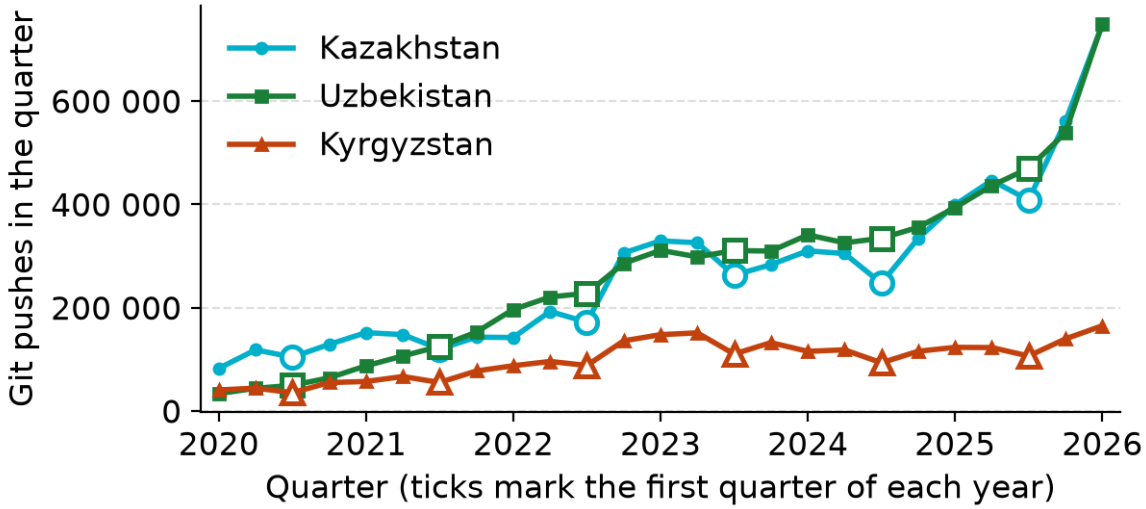


Figure 3 - Git pushes by quarter; open symbols mark the third quarters (source: GitHub Innovation Graph [1])

In Kazakhstan and Uzbekistan the check is passed. Yearly git pushes grew 4.19 times in Kazakhstan between 2020 and 2025 while accounts grew 4.29 times, and 9.65 times in Uzbekistan while accounts grew 8.49 times. Pushes per account per year therefore stayed almost flat in Kazakhstan (5.24 in 2020, 5.04 in 2025) and rose slightly in Uzbekistan (from 5.86 to 6.18). Uzbekistan first had more pushes than Kazakhstan in 2021 Q3 and has had more repositories than Kazakhstan in every quarter since 2023 Q1. In 2025 the two countries made almost the same number of pushes (1,813,424 and 1,837,684), and in 2026 Q1 the difference was only 1,626 pushes.

In Kyrgyzstan the check fails. Its accounts grew 5.21 times between the end of 2020 and the end of 2025, but its yearly pushes grew only 2.82 times, so pushes per account fell by almost half, from 9.92 to 5.11. In Kyrgyzstan, a large part of the growth in accounts was not matched by more uploaded code.

Languages: Python grew faster than the developer base

Figure 4 shows the number of developers who pushed to Python repositories in each quarter. In Kazakhstan the yearly average rose from 1,452 in 2020 to 7,419 in 2025 (5.11 times), faster than accounts (4.29 times) and pushes (4.19 times). In Uzbekistan it rose from 453 to 6,407 (14.15 times), again faster than accounts (8.49 times) and pushes (9.65 times). In Kyrgyzstan Python users grew more slowly than accounts (4.11 against 5.21 times) but faster than pushes (2.82 times).

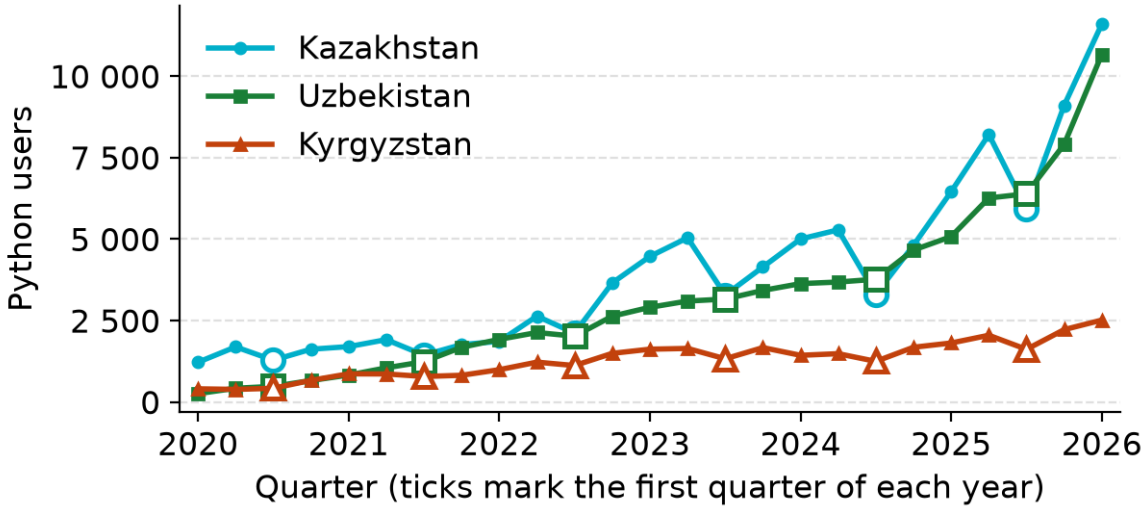


Figure 4 - Developers who pushed to Python repositories, by quarter; open symbols mark the third quarters (source: GitHub Innovation Graph [1]; first chart for Kazakhstan and Uzbekistan sent by the author, Kyrgyzstan added)

Table 3 places Python among the ten languages with the most users in Kazakhstan in 2025. The languages did not grow at the same pace. In Kazakhstan the web trio HTML, CSS and JavaScript grew about 3.5 times, Java and Shell 2.8 times and Python 5.1 times, while TypeScript (8.1 times) and Dockerfile (8.7 times) grew fastest. Every language in the table for which the growth in Uzbekistan can be computed grew faster there than in Kazakhstan, and TypeScript in Uzbekistan grew more than 29.6 times. By 2025 the six most used languages were the same in both countries and in the same order: HTML, CSS, JavaScript, Python, TypeScript and Dockerfile. In 2020 the lists had differed: Kazakhstan had Java in fifth place, and Uzbekistan had PHP in fourth place, ahead of Python.

Table 3 - Developers who pushed to repositories in each language, yearly mean per quarter, 2020 and 2025

Language	KZ 2020	KZ 2025	KZ growth, times	UZ 2020	UZ 2025	UZ growth, times
HTML	3,542	12,228	3.45	2,033	13,465	6.62
CSS	2,660	9,389	3.53	1,553	11,194	7.21
JavaScript	2,566	9,076	3.54	1,422	10,433	7.34
Python	1,452	7,419	5.11	453	6,407	14.15
TypeScript	452	3,675	8.13	127*	3,754	more than 29.56*
Dockerfile	340	2,958	8.71	132	1,984	15.06
Java	973	2,711	2.79	293	1,111	3.79
Shell	891	2,488	2.79	451	1,752	3.89

C++	339	1,356	4.00	137*	1,106	more than 8.07*
Go	198	1,340	6.76	not published	376	not available

Note. Source: GitHub Innovation Graph [1]; calculations for this study. The ten languages with the most users in Kazakhstan in 2025. * In 2020 the number for Uzbekistan was published for only some quarters (TypeScript three, C++ one), because in the other quarters fewer than 100 developers used the language; the mean of the published quarters is shown, and the true growth is larger than the number given. Go in Uzbekistan was below this threshold in all quarters of 2020.

The third-quarter dip (exploratory)

The Python chart has a clear pattern: Kazakhstan's line drops in every third quarter (July to September) and recovers in the fourth. The git push charts for Kazakhstan on the Innovation Graph website show the same drop for all code uploads. Table 4 measures it.

Table 4 - Third-quarter dip, 2020-2025: the third quarter compared with the midpoint of the second and fourth quarters

Country and series	Years with Q3 below Q2	Years with a dip	Mean dip, %	t	p	p (Holm)
KZ, git pushes	6 of 6	6 of 6	−19.6	−7.39	0.0007	0.005
KZ, Python users	6 of 6	6 of 6	−28.5	−11.32	0.0001	0.0008
KZ, developer accounts	0 of 6	6 of 6	−1.1	−2.35	0.065	0.20
UZ, git pushes	0 of 6	5 of 6	−3.5	−2.08	0.092	0.20
UZ, Python users	1 of 6	6 of 6	−8.7	−5.28	0.003	0.016
UZ, developer accounts	0 of 6	5 of 6	−0.6	−2.08	0.092	0.20
KG, git pushes	6 of 6	6 of 6	−22.6	−14.61	< 0.001	< 0.001
KG, Python users	5 of 6	6 of 6	−17.9	−6.86	0.001	0.006
KG, developer accounts	0 of 6	6 of 6	−1.2	−2.64	0.046	0.18

Note. Source: GitHub Innovation Graph [1]; calculations for this study. "Years with a dip" counts years in which the third quarter was below the midpoint of its two neighbours on a logarithmic scale; t and p are from a two-sided one-sample t-test of the six yearly values against zero (5 degrees of freedom). The p-values are descriptive, because the test was chosen after the dip had been seen; the last column adjusts them for the nine tests in this table (Holm method).

In Kazakhstan the third quarter was below the second in every year from 2020 to 2025, both for git pushes and for Python users. On average, git pushes in the third quarter were 19.6% below the midpoint of the neighbouring quarters, and Python users 28.5% below; both stay clearly below 0.05 after the Holm adjustment. Without 2022, when the third quarter was followed by an unusual jump (see below), the mean dip in pushes is 17.6%. Part of this contrast comes from a spring peak: against a centred four-quarter moving average, which spreads one high neighbour over a whole year, the second quarter in Kazakhstan (2021-2025) is 6.8% above trend for pushes and 18.6% for Python users, and the third quarter is 15.2% below trend for

pushes and 22.0% for Python users. Kyrgyzstan shows a dip of similar size in pushes (22.6% below the midpoint, 17.5% below trend). In Uzbekistan the dip in pushes is small (3.5%) and not statistically significant, and pushes never fell from the second to the third quarter; a smaller but regular dip appears in Python users (8.7%, Holm-adjusted p = 0.016). The number of accounts barely reacts to the season (dips of about 1%, not significant after adjustment), which is expected for a count that includes inactive accounts.

Is this dip special to Kazakhstan? The comparison of 146 economies (Figure 5) says no. Across all of them the median third-quarter value of git pushes was 5.3% below the midpoint of its neighbours, and 63.0% of economies had a dip. The dip is a Northern Hemisphere pattern: the median was −6.5% for the 120 northern economies and +4.6% for the 26 southern ones (Mann-Whitney U = 863, p = 0.0004), although 9 of the 26 southern economies still dip. Most southern capitals in the data lie in the tropics, where the seasons differ little; outside the tropics the contrast is sharper, −11.9% for 76 northern economies against +4.8% for 9 southern ones (p = 0.003). The mirror test points the same way: outside the tropics the first quarter, which is summer in the south, is below its neighbours in the southern economies (−6.9%) but not in the northern ones (+1.3%, p = 0.008), while across all economies this first-quarter difference is not significant (p = 0.45). Kazakhstan's dip is the 15th deepest of the 146; in the World Bank's Europe and Central Asia region (47 economies, median third-quarter dip 15.7%) Kazakhstan ranks 12th, next to Poland (19.7%) and Finland (19.6%), and Kyrgyzstan 4th; the unusual country is Uzbekistan, which ranks 46th of 47. Across all 146 economies, faster growth in pushes from 2020 to 2025 went together with a smaller dip (Spearman ρ = 0.55, p < 0.001; ρ = 0.59 among the northern economies alone), and Uzbekistan had the 9th fastest growth in pushes of the 146.

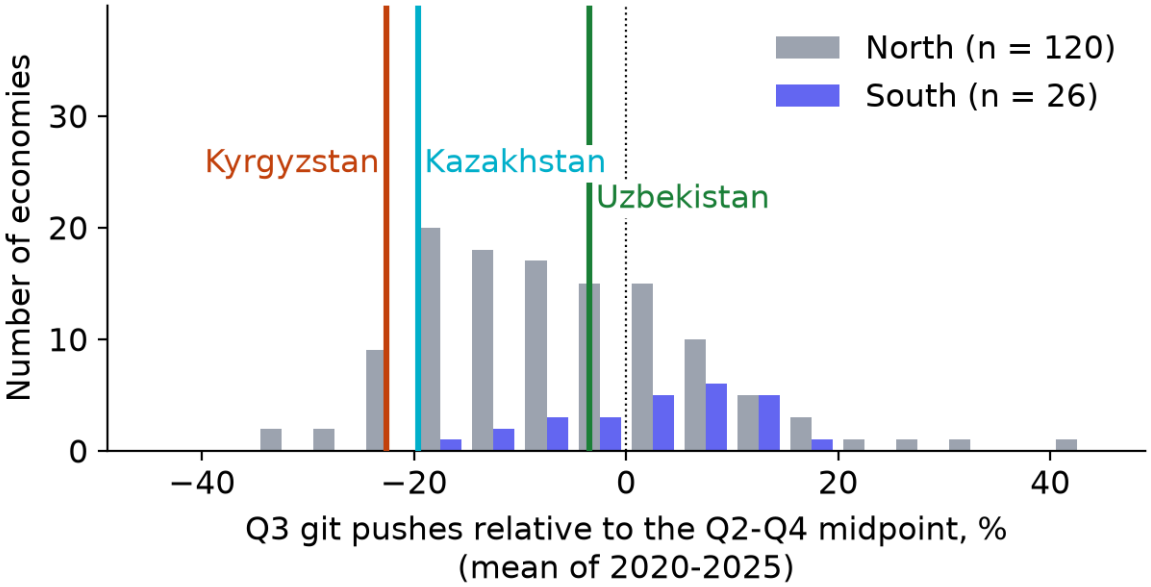


Figure 5 - Third-quarter dip in git pushes in 146 economies, mean of 2020-2025, by hemisphere of the capital city (source: GitHub Innovation Graph [1], World Bank [9]; calculations for this study)

Two unusual quarters (exploratory)

In 2022 Q4 developer accounts grew by 13.5% in Kazakhstan and by 15.8% in Kyrgyzstan in one quarter. For both countries this was the fastest quarter of 2021-2023 (the next fastest were 9.2% and 11.8%), and it was 8.0 and 10.3 percentage points above the growth of

all economies together (5.5%). Fourth quarters are often strong in both countries, however: Kazakhstan's accounts also grew by 11.5% in 2020 Q4 and 12.1% in 2025 Q4, and Kyrgyzstan's by 14.1% and 12.2%. In Uzbekistan the 2022 Q4 growth (14.1%) was in line with its own pace in 2021 and 2022 and ranks only 7th of its twelve quarters in 2021-2023. In the same quarter git pushes rose by 78.1% in Kazakhstan, 70.3 percentage points above all economies together and the largest quarterly change in its whole series, and by 54.3% in Kyrgyzstan, within the range of its other fourth quarters (20.0-55.7%); Kazakhstan's other fourth quarters rose by 7.8-37.3%, so only part of its rise is the usual recovery after the third-quarter dip. In 2026 Q1 git pushes rose by 33.4% in Kazakhstan and 39.1% in Uzbekistan, but the total of all economies rose by 29.6% and the median economy by 31.7%, so this last jump is a global event rather than a Central Asian one.

Discussion

The answer to the research question is that Uzbekistan grows faster and has nearly reached Kazakhstan in absolute numbers, while Kazakhstan still has 2.1 times more developer accounts per resident, and that languages do not grow at the same pace. The planned check matters for how these numbers should be read: in Kazakhstan and Uzbekistan code activity grew together with accounts, so the growth is not just empty registrations. Kyrgyzstan is different: its accounts grew faster than its code, and the average account made about half as many pushes in 2025 as in 2020. The data cannot say why.

The third-quarter dip is the clearest pattern in Kazakhstan's data, and the comparison with other economies changes how it should be read. It is not specific to Kazakhstan: it is typical of the Northern Hemisphere, where the main summer holidays fall in July and August. In the Southern Hemisphere it is much weaker or reversed, and outside the tropics, where July to September is winter in the south, the southern economies show a dip in their own summer quarter instead. This fits GitHub's own observation that activity changes in holiday seasons and when students are out of school [12]. The dip in Kazakhstan is deeper for Python users than for all pushes, which would be consistent with Python users including many students, but the Innovation Graph does not show who the developers are, so this remains an interpretation. The more surprising result is that Uzbekistan has only a small summer dip. Across economies, fast growth goes together with weaker dips, so one explanation is that so many new developers join Uzbekistan's community every quarter that the summer drop is partly hidden; this is a correlation, not a proof.

The fast growth of accounts in Kazakhstan and Kyrgyzstan in 2022 Q4 coincides in time with the relocation of highly active developers from Russia and Belarus documented by Wachs, who estimated arrivals equal to 13% of Kazakhstan's and 7% of Kyrgyzstan's highly active developers [6]. Because both countries also had strong fourth quarters in 2020 and 2025, this is a coincidence in time, not a test of cause, and it is not visible in Uzbekistan's data, although Wachs's estimates for Uzbekistan and Kyrgyzstan were similar (8% and 7%, based on 15 and 9 developers).

Limitations. The Innovation Graph covers only public activity on GitHub and only counts above the threshold of 100 developers, and because location is assigned by IP address, VPN use and travel can move activity between countries [2]. The third-quarter test uses six years, enough to show a regular pattern but not to study how it changes over time. Population for 2026 Q1 uses the 2025 value. The seasonal and cross-economy analyses were added after the dip was noticed and should be read as exploratory. Finally, the growth on GitHub coincides

with growing official IT export figures in both countries [10, 11], but this paper does not test whether more developers lead to more exports.

Conclusion

Between 2020 and 2026 the number of GitHub developer accounts grew 6.7 times in Kazakhstan, 8.4 times in Kyrgyzstan and 16.1 times in Uzbekistan. Uzbekistan has nearly caught up with Kazakhstan in absolute numbers, but Kazakhstan still has about twice as many developers per resident. In Kazakhstan and Uzbekistan the growth in accounts is backed by a similar growth in code, which is not the case in Kyrgyzstan. Developers pushing to Python repositories grew faster than the developer base in both Kazakhstan and Uzbekistan. Kazakhstan's activity drops every summer, by about a fifth against the neighbouring quarters and about 15% against trend, but the exploratory comparison of 146 economies suggests it is a seasonal pattern shared by most Northern Hemisphere economies, deep but typical for Europe and Central Asia; the exception that needs explaining is Uzbekistan, where the summer dip is much smaller.

Author contributions and use of AI tools

Yerganat Mukhamedzhan chose this topic from five research directions proposed by his mentor, agreed the research question, and proposed Kyrgyzstan as the third country. He produced the first version of Figure 1 (Kazakhstan and Uzbekistan) by running, on the GitHub Innovation Graph data, a script that was written by an AI assistant (Claude, Anthropic); his mentor changed its file paths so that it ran on his computer. He also sent the first chart of Python users for Kazakhstan and Uzbekistan, from a language file supplied by his mentor. He found and summarised in Russian, with the help of an AI assistant, three of the studies reviewed here [3, 4, 5], and he viewed and saved the git push charts for Kazakhstan and Uzbekistan on the Innovation Graph website. The third-quarter dip was first noticed by his mentor in the author's Python chart; the push charts he saved later showed the same drop. The data sources and the other studies cited were found by his mentor. His research mentor proposed the five directions, chose the target journal and wrote the analysis plan, which the author approved. The final data processing, the statistical tests, the exploratory analyses, the final figures and tables, the English text of the manuscript (including translation of the author's summaries) and the checking of references were carried out by the mentor with the help of an AI assistant (Claude, Anthropic).

Data and code availability

The GitHub Innovation Graph is openly available at https://github.com/github/innovationgraph under the CC0 licence; population data are from the World Bank. The analysis scripts, the exact data files used and all result tables are kept by the author's research mentor and are available on request.

References:

1. GitHub. GitHub Innovation Graph [Electronic resource]: dataset, version 1.0.11. – 2026. – URL: https://github.com/github/innovationgraph (accessed: 11.10.2026).

2. GitHub. GitHub Innovation Graph Metrics: datasheet [Electronic resource]. – 2026. – URL: https://github.com/github/innovationgraph/blob/main/docs/datasheet.md (accessed: 11.10.2026).

3. Wachs J., Nitecki M., Schueller W., Polleres A. The geography of open source software: evidence from GitHub // Technological Forecasting and Social Change. – 2022. – Vol. 176. – Art. 121478. – DOI: 10.1016/j.techfore.2022.121478.

4. Juhász S., Wachs J., Kaminski J., Hidalgo C.A. The software complexity of nations // Research Policy. – 2026. – Vol. 55, No. 3. – Art. 105422. – DOI: 10.1016/j.respol.2026.105422. Supplementary information in the arXiv version: arXiv:2407.13880v2. – URL: https://arxiv.org/abs/2407.13880 (accessed: 11.10.2026).

5. Quispe A., Grijalba R. Impact of the availability of ChatGPT on software development: a synthetic difference in differences estimation using GitHub data [Electronic resource] // arXiv. – 2024. – arXiv:2406.11046. – URL: https://arxiv.org/abs/2406.11046 (accessed: 11.10.2026).

6. Wachs J. Digital traces of brain drain: developers during the Russian invasion of Ukraine // EPJ Data Science. – 2023. – Vol. 12. – Art. 14. – DOI: 10.1140/epjds/s13688-023-00389-3.

7. Mussin A., Shorakhan B., Bakhyt A., Kassymkhan M., Karimov T. Comparison of inflation and unemployment between Kazakhstan and Uzbekistan // Central Asian Scientific Journal. – 2026. – No. 2(30), Vol. 1. – P. 472–482.

8. Yanson D., Dautov D., Abdul-ghani M., Nazar S., Tsoy L., Kalambay A. The transmission mechanism of value-added tax and its impact on consumer price index: comparative analysis of Kazakhstan and Uzbekistan // Central Asian Scientific Journal. – 2026. – No. 2(30), Vol. 1. – P. 512–517.

9. World Bank. Population, total (SP.POP.TOTL) [Electronic resource] // World Development Indicators. – 2026. – URL: https://data.worldbank.org/indicator/SP.POP.TOTL (accessed: 11.10.2026).

10. National Statistics Committee of the Republic of Uzbekistan. Service export volume of IT Park participants (indicator 1.08.02.0045) [Electronic resource]. – 2026. – URL: https://siat.stat.uz/data/3358/ (accessed: 11.10.2026).

11. Ministry of Artificial Intelligence and Digital Development of the Republic of Kazakhstan. Kazakhstan becomes a net technology exporter: IT service exports exceed $1.1 billion in 2025 [Electronic resource]. – 10.04.2026. – URL: https://www.gov.kz/memleket/entities/maidd/press/news/details/1198462?lang=en (accessed: 11.10.2026).

12. Elliott T. The State of the Octoverse: breaks and holidays [Electronic resource] // The GitHub Blog. – 2018. – URL: https://github.blog/news-insights/the-library/breaks-and-holidays/ (accessed: 11.10.2026).

ПӘНДЕР АРАЛЫҚ ҒЫЛЫМДАР – МЕЖДИСЦИПЛИНАРНЫЕ НАУКИ – INTERDISCIPLINARY SCIENCES

UDC 81'272:004.738.5

Albiyev Asser

Grade 11 student
Quantum STEM School
(Astana, Kazakhstan)

HOW MUCH OF AN ARTICLE COUNT IS HUMAN WORK? DECOMPOSING THE CREATION OF FIVE WIKIPEDIAS, 2010–2025

Abstract: The number of articles in a language's Wikipedia is routinely quoted as evidence of that language's presence online, including in a 2026 study of Kazakh digital language policy that reports 241,858 articles for the Kazakh edition. An article count is a count of pages, not of human work, and pages can be created automatically. Using the Wikimedia Analytics API, this study decomposes every content page created in the Kazakh, Uzbek, Kyrgyz, Azerbaijani, Tatar and Russian editions between January 2010 and August 2025 by the class of account that created it. Accounts carrying the platform's bot flag created 50.4 per cent of the Kazakh pages, 64.1 per cent of Kyrgyz and 90.0 per cent of Tatar, against 0.3 per cent of Azerbaijani. The flag is then shown to be insufficient. A burst rule — a month in which human-flagged page creation is at least ten times its own trailing median — identifies four Kazakh months, thirteen Uzbek and ten Kyrgyz, and none at all in Azerbaijani, Tatar or Russian. In those months a newly created page receives 2.5 to 3.8 subsequent edits against 17.1 to 21.8 in ordinary months, and carries roughly half the net bytes, which is the signature of automated stub creation by accounts the flag does not catch. Treating the burst excess as automation lowers the human share of Kazakh creation from 48.3 to 27.2 per cent. The same rule run as a placebo on already-flagged bot creation recovers 87 to 97 per cent of it in the three editions where it fires, so the signature it detects is automation rather than volume. Ranking the five non-Russian editions by articles gives Tatar first and Azerbaijani fourth; ranking them by human-created pages reverses this, and Azerbaijani leads on both corrected and uncorrected counts. Two findings are specific to Kazakh: no flagged bot has created an article there since 2012, and the human share of edits rose from 17.2 to 95.1 per cent over the period, but this rise is a collapse in bot activity rather than growth in human activity, and the core editing community averaged 44.2 people a month during the 2011 campaign against 37.1 in 2024. An article count should not be read as a measure of human work without this decomposition, and the platform's own bot flag is not sufficient to perform it.

Keywords: Wikipedia; Kazakh language; bots; automated content; digital language vitality; Central Asia; article count

1. Introduction

Wikipedia keeps a public counter of how many articles each of its language editions contains, and that counter has become a shorthand for how well a language is doing online. The habit reaches into the research literature. A 2026 study of Kazakh digital linguistic sovereignty lists the Kazakh Wikipedia among its indicators of the language's digital presence, reporting

that "the section comprised 241,858 articles covering a wide range of subjects, with approximately 2 million edits and a membership of over 10,000 registered participants" [1]. The count is used there exactly as it is used in public argument: as a quantity of Kazakh-language knowledge that exists.

The counter does not measure that. It measures pages. A page can be written by a person over several sittings, or generated by a program from a database in a fraction of a second, and the counter treats the two identically. Where automated creation has been heavy this is not a small correction. The Cebuano edition of Wikipedia is the extreme case: it is the largest non-English edition by article count, and almost the whole of it was produced by a single article-creating program [11].

Kazakhstan has a specific reason to care. In 2011 the foundation WikiBilim ran a campaign to grow the Kazakh edition, and the printed Kazakh National Encyclopedia donated its content to it [12]; a Wikimedia Foundation post from July 2012 records that "Today we have more than 130,000 articles and more than 200 active editors" [12]. Fifteen years later that campaign is still inside the counter that gets quoted. Nobody has said how much of it.

This paper asks a narrow, answerable question: of the content pages created in a language edition between 2010 and 2025, what share was created by a human being? We ask it for Kazakh and for four comparison editions — Uzbek, Kyrgyz and Azerbaijani, which face broadly similar circumstances, and Tatar, which is known for heavy automated creation — with Russian as a large reference.

The first answer is the obvious one, and it is already substantial: on the platform's own bot flag, half of the Kazakh pages created since 2010 were created by bots. The second answer is the one we think matters more. The flag is not sufficient. Mass creation also runs through ordinary, unflagged accounts, and it leaves a measurable trace that the flag misses. Correcting for that trace roughly halves the human share again, and it changes which edition sits at the top of the table.

We report one further result that cuts against a simple pessimistic reading. Kazakh bot creation stopped in 2012 and has not resumed, and the human share of Kazakh editing has risen steadily since. We also show why that rise is weaker evidence than it looks.

2. Background and related work

Automated accounts have been part of Wikipedia almost since its beginning, and the research on them is mostly about what they do rather than how much of the encyclopedia they built. Geiger describes bots as socio-technical actors embedded in the community's rules rather than as mere tools [7]. Halfaker and Riedl distinguish bots proper from "cyborgs" — human editors working through semi-automated tools that combine computational speed with human judgement [6]. That distinction is central to this paper: a cyborg edits from a human account, and no bot flag is attached to it. Tsvetkova and colleagues show that bots interact with each other in patterned ways over long periods [5], and Zheng and colleagues build a taxonomy of bot functions on the English edition and classify 1,601 of them [8], establishing that "bot" is not one kind of thing.

The work closest to ours in method is Steiner's, which monitors real-time edit activity across all language editions and reports that Wikipedias and Wikidata together are edited roughly 50 per cent by bots and 23 per cent by anonymous users [3, 4]. Steiner measures edits. We measure page creation, which is the quantity the article counter actually accumulates, and we measure it per edition over fifteen years rather than in a live window.

Reeves and Simperl review the literature on automated content generation for Wikipedia [9]. It is a review of generation techniques and evaluation methods; it does not provide a per-edition accounting of how many existing articles were machine-made, which is the gap this paper addresses.

On the other side sits the literature that treats online presence as a measure of a language's health. Kornai argues that most of the world's languages will not make the transition to digital use at all, and that the digital domain is where language vitality is now decided [2]. That framing is what gives an article count its rhetorical force, and it is also what makes an uncorrected count dangerous: if the count is the vitality measure, then a language can appear to be thriving on the strength of work no speaker did.

To our knowledge no study decomposes the article creation of the Central Asian Turkic Wikipedia editions by the class of account responsible, and none tests whether the platform's own bot flag is adequate for such a decomposition. Both are what we do here.

3. Data

All series come from the Wikimedia Analytics REST API [10], which is public, requires no account and no key, and exposes monthly aggregates of the MediaWiki history. We use six editions — kk, uz, ky, az, tt and ru — over the 188 months from January 2010 to August 2025, the last month available at the time of the pull on 23 September 2026.

For each edition and each month we pulled four series, each broken out by the four editor classes the API distinguishes (`user`, `group-bot`, `name-bot`, `anonymous`): content-namespace edits, newly created content pages, distinct editors, and net bytes added. We also pulled human editors split by monthly activity level, and the live article count from each edition's own MediaWiki statistics endpoint. The API returns only months in which a series is non-zero; an absent month is read as zero, which matters for bot series that run in short campaigns.

Two properties of the classes should be stated plainly. `group-bot` means the account held the bot user group; `name-bot` means the account's name contains "bot" without that group; `anonymous` means an unregistered editor. The classification is a property of the account, not of the behaviour, and Section 5.2 is about what that costs.

The window covers most but not all of each edition's stock. Pages created in the window number 97 per cent of the live Kazakh article count, 99 per cent of Kyrgyz, 85 per cent of Azerbaijani, 81 per cent of Uzbek, 75 per cent of Russian and 70 per cent of Tatar. The residual is pages created before 2010 net of pages deleted since. We therefore state our primary results as a decomposition of the 2010–2025 creation flow, which is fully measured, and treat any statement about the present stock as an extrapolation that carries the coverage ratio with it.

4. Method

The first step is arithmetic: for each edition, sum created content pages over the window by editor class and express each class as a share.

The second step is the substantive one. If mass creation also runs through unflagged human accounts, the class shares understate automation, and the understatement will be worst exactly where the campaigns were largest. We need a way to detect automation from behaviour rather than from account status.

We use a burst rule. For each edition and each month from the thirteenth onward, compute the median of human-flagged page creation over the preceding twelve months. A month is a creation burst when human-flagged creation is at least *K* times that trailing median and also clears an absolute floor of 300 pages, which prevents the rule from firing on the ordinary noise

of a very small edition. The burst excess is the month's creation minus its own trailing median — that is, the part that is not explained by the edition's own recent human pace. We treat the burst excess as automation. Because a genuine human surge would also be caught, the excess is an upper bound on unflagged automated creation, and the flagged-bot count is a lower bound; the truth lies between, and we report both.

We set *K* = 10 for the main results and sweep *K* over 5, 10, 20 and 50, and the floor over 100, 300 and 1,000, so that no conclusion rests on a chosen threshold.

A rule of this kind must be shown to detect automation rather than simply volume, so we subject it to three checks. First, two behavioural corroborators computed independently of the rule: the number of subsequent edits a newly created page receives, and the net bytes added per new page, both compared between burst and ordinary months. Automated stub creation should produce pages that are barely revisited and short. Second, a placebo: the identical rule applied to creation that is *already* flagged as bot. If the burst shape is the automation signature, it should recover most flagged-bot creation; if it merely flags large months, it proves nothing. Third, the editions where we expect no unflagged campaigns — Azerbaijani, Russian — function as negative controls, and the rule is expected to stay silent there.

Finally, for the question of whether the human share is growing, we aggregate to years and fit ordinary least squares of the annual human share of edits on the year, reporting the slope, its *t* statistic and the fit. Sixteen annual points is a small sample and we treat weak *t* statistics as weak evidence rather than as evidence of absence.

5. Results

5.1 What the article counts are made of

Table 1 — Composition of content-page creation, January 2010 to August 2025.

"Bot" is the platform's own flag (group-bot plus name-bot). Core human editors is the mean monthly count of accounts making 25 or more content edits, over the last twelve months of the window. Articles per core editor divides the live article count by that mean.

Edition	Live articles	Created 2010–2025	Coverage	Bot share of creation	Bot share of edits	Core human editors / month	Articles per core editor
Kazakh (kk)	245,273	238,743	97%	50.4%	49.9%	34	7,232
Uzbek (uz)	361,810	292,212	81%	38.7%	57.3%	58	6,211
Kyrgyz (ky)	76,549	75,853	99%	64.1%	36.7%	7	11,067
Azerbaijani (az)	217,531	185,357	85%	0.3%	25.1%	73	2,973
Tatar (tt)	709,127	499,864	70%	90.0%	87.1%	8	94,550
Russian (ru)	2,118,986	1,589,009	75%	3.6%	19.3%	1,071	1,979

Half of everything the Kazakh edition created since 2010 was created by flagged bots: 120,333 pages against 115,427 from human-flagged accounts and 2,983 from unregistered editors. Kyrgyz is higher at 64.1 per cent and Tatar higher still at 90.0 per cent. Uzbek, at 38.7 per cent, looks the healthiest of the four.

The outlier runs the other way. The Azerbaijani edition has 217,531 articles, within eleven per cent of Kazakh, and flagged bots created 533 of the 185,357 pages it added in the window — three tenths of one per cent. Two editions of comparable size, built in almost opposite ways.

The last column is the crudest possible sanity check and it is already informative. Tatar carries 94,550 live articles for every core human editor active in the last year. Kazakh carries 7,232, Azerbaijani 2,973, Russian 1,979. An encyclopedia is maintained, not just created, and these ratios describe how much material sits behind each person available to maintain it.

5.2 The flag is not enough

Table 1 rests entirely on account status. Section 2 gives the reason to distrust it: semi-automated editing from ordinary accounts is a recognised category [6], and it carries no flag.

The Kazakh series shows the problem immediately. Human-flagged accounts created a median of 271 pages a month across the window. In June 2011 they created 18,448, in July 21,972, in August 9,456. Those three months are 229, 243 and 63 times the trailing twelve-month median, and they alone account for 49,876 pages.

Table 2 — Burst detection on human-flagged page creation, K = 10, floor = 300 pages.

Burst excess is creation above the month's own trailing twelve-month median. The last two columns are computed independently of the rule.

Edition	Burst months	Burst excess pages	Excess as share of human creation	Human share, flag only	Human share, corrected	Later edits per new page, burst	Later edits per new page, ordinary
Kazakh (kk)	4	50,498	43.7%	48.3%	27.2%	2.8	17.1
Uzbek (uz)	13	92,530	52.1%	60.7%	29.1%	3.8	19.6
Kyrgyz (ky)	10	13,755	52.3%	34.7%	16.6%	2.5	21.8
Azerbaijani (az)	0	0	0.0%	97.8%	97.8%	—	21.9
Tatar (tt)	0	0	0.0%	9.7%	9.7%	—	10.9
Russian (ru)	0	0	0.0%	89.7%	89.7%	—	39.5

The rule fires in three editions and stays silent in three. Where it fires, it fires hard: the excess is 43.7 per cent of all human-flagged Kazakh creation, 52.1 per cent of Uzbek and 52.3 per cent of Kyrgyz. Treating that excess as automation cuts the human share of Kazakh creation from 48.3 to 27.2 per cent, of Uzbek from 60.7 to 29.1 and of Kyrgyz from 34.7 to 16.6.

The last two columns are the reason to believe this is not an artefact of the rule. They are not part of the rule and could have come out flat. In Kazakh burst months a newly created page goes on to receive 2.8 later edits; in ordinary Kazakh months, 17.1. The same contrast appears in Uzbek (3.8 against 19.6) and Kyrgyz (2.5 against 21.8). Pages created in burst months are made once and left. Net bytes point the same way, if less sharply: 2,950 bytes per new page in Kazakh burst months against 6,953 in ordinary ones, 2,870 against 7,679 in Kyrgyz, 3,745 against 4,738 in Uzbek.

The Uzbek bursts are recent and sustained — eight months across 2022, and June 2024 alone with 37,456 pages against a trailing median of 1,174. The Kazakh bursts are all in 2011,

and they coincide with the documented WikiBilim campaign and the Kazakh National Encyclopedia donation of that year [12]. We take that coincidence as corroboration that the rule detects a real event, not as an identification of who performed it, which the aggregate data cannot establish.

5.3 Does the detector point the right way?

Table 3 — Placebo. The identical burst rule applied to creation that already carries the bot flag.

A rule that detects automation should recover most of it; a rule that merely detects large months would not.

Edition	Pages created by flagged bots	Falling in months the rule flags	Share recovered
Kazakh (kk)	120,333	105,291	87%
Uzbek (uz)	113,174	51,456	45%
Kyrgyz (ky)	48,555	47,319	97%
Azerbaijani (az)	248	0	0%
Tatar (tt)	440,707	389,384	88%
Russian (ru)	15,121	7,161	47%

Run against creation we already know to be automated, the rule recovers 87 per cent of it in Kazakh, 97 per cent in Kyrgyz and 88 per cent in Tatar. It recovers nothing in Azerbaijani, where flagged bots created only 248 pages in fifteen years and never enough in one month to clear the floor — the correct answer, not a failure. Uzbek and Russian sit near half, meaning their flagged bot work is more evenly spread and less campaign-shaped; where that is so, our correction is conservative, since steady automation does not burst.

Taken with the two behavioural corroborators in Table 2, this is the evidence that the burst shape is the automation shape. It is not proof that every flagged month was automated, and we do not claim it: the excess remains an upper bound.

Table 4 — Corrected human share of content-page creation under different burst thresholds K, floor held at 300.

Varying the floor over 100, 300 and 1,000 pages moves Kazakh not at all (27.2% throughout), Uzbek from 29.0% to 29.8% and Kyrgyz from 16.0% to 20.2%.

K	Kazakh	Uzbek	Kyrgyz	Azerbaijani	Tatar	Russian
5	24.8%	24.8%	15.6%	87.4%	9.7%	89.7%
10	27.2%	29.1%	16.6%	97.8%	9.7%	89.7%
20	27.6%	38.4%	17.9%	97.8%	9.7%	89.7%
50	27.6%	56.8%	28.4%	97.8%	9.7%	89.7%

The sensitivity sweep separates a robust conclusion from a fragile one, and we report both. The Kazakh correction is robust: across a tenfold range of *K* the corrected human share moves only between 24.8 and 27.6 per cent, and changing the floor does not move it at all. The Uzbek correction is not: at *K* = 5 it gives 24.8 per cent and at *K* = 50 it gives 56.8 per cent. The Uzbek campaigns are long enough to raise their own trailing median, so how much of them the rule catches depends on where the threshold sits. We therefore claim a level for Kazakh and only a direction for Uzbek. Kyrgyz sits in between, ranging from 15.6 to 28.4 per cent, and Azerbaijani flips only at the loosest threshold, where a single 2020 month is caught.

5.4 The league table, three ways

Table 5 — Five editions ranked by the headline article count and by human-created pages in the 2010–2025 window, before and after the burst correction. Russian excluded as the reference case.

Rank	By article count	By human-created pages (flag only)	By human-created pages (corrected)
1	Tatar — 709,127	Azerbaijani — 181,308	Azerbaijani — 181,308
2	Uzbek — 361,810	Uzbek — 177,468	Uzbek — 84,938
3	Kazakh — 245,273	Kazakh — 115,427	Kazakh — 64,930
4	Azerbaijani — 217,531	Tatar — 48,525	Tatar — 48,525
5	Kyrgyz — 76,549	Kyrgyz — 26,312	Kyrgyz — 12,557

Ranking by the public counter puts Tatar first and Azerbaijani last but one. Ranking by pages a person created inverts that: Azerbaijani first, Tatar fourth, and the gap between first and fourth is a factor of 3.7 rather than the factor of 3.3 by which Tatar appeared to lead. Azerbaijani leads on both the uncorrected and the corrected count, and it is the only edition whose position is unchanged by the correction — because there is nothing there to correct.

For Kazakh the two counts differ by a factor of 3.8: 245,273 articles, and 64,930 pages created by a human being under the corrected reading — 115,427 if only the flag is trusted. Both figures are for the 2010–2025 window; the live stock includes pre-2010 pages, which for Kazakh is a small residual given the 97 per cent coverage.

5.5 Is the human share growing?

Table 6 — Annual human share of content edits regressed on year, 2010–2025, sixteen annual observations. 2025 covers January to August only.

Edition	Slope, pp per year	t	R^2	Human share of edits, 2010	Human share of edits, 2025
Kazakh (kk)	+4.80	4.21	0.56	17.2%	95.1%
Uzbek (uz)	+4.45	2.94	0.38	2.3%	75.2%
Kyrgyz (ky)	+2.03	1.81	0.19	29.8%	86.9%
Azerbaijani (az)	+1.86	2.00	0.22	55.9%	83.6%
Tatar (tt)	−1.55	−1.34	0.11	26.3%	17.8%
Russian (ru)	−0.11	−0.30	0.01	60.1%	46.0%

On edits the Kazakh trend is the strongest in the set: +4.80 percentage points a year, t = 4.21, and a rise from 17.2 per cent in 2010 to 95.1 per cent in the first eight months of 2025. Taken alone this reads as a decisively human edition.

The annual series says why that reading is too generous. Kazakh bot edits fell from 278,875 in 2011 and 436,448 in 2012 to 2,980 in 2025. Kazakh human edits went 5,422 in 2010, 204,444 in 2011, 80,828 in 2016, 102,969 in 2024, and 57,687 in the eight months of 2025. The human share rose overwhelmingly because the denominator collapsed. Human editing has not grown since the middle of the decade; the bots left.

The same point holds for creation, in Kazakh's favour. No flagged bot has created a Kazakh content page since 2012: the 120,333 bot-created pages are 49,468 from 2011 and

70,865 from 2012 and nothing after. Every Kazakh article created in the last thirteen years was created from a human-flagged account, and no month since 2011 trips the burst rule. Whatever the stock is made of, the recent flow is human.

The size of the community that produces that flow is the sobering part. The mean monthly count of Kazakh accounts making 25 or more content edits was 4.0 in 2010, 44.2 in the campaign year of 2011, fell to 12.8 by 2017, and recovered to 37.1 in 2024 and 36.0 in the first eight months of 2025. We fitted both a linear and a quadratic trend to these sixteen points and neither is distinguishable from noise ($t = 1.45$ and $t = 1.61$ respectively), so we report no trend and state the levels instead. The level is the finding: fourteen years after the campaign, the core Kazakh editing community has not returned to the size it reached during it.

The following is the author's own account, written before seeing the results of this study (Telegram, 20 and 23 September 2026), lightly edited and translated.

I open the English Wikipedia for almost everything. It has a large community that edits often, so it is usually the most current, and its articles are the most complete. My schoolwork is in English too, so if I need to quote something it is easiest to take it from there. My classmates do the same, except the ones whose English is weak; they use the Russian Wikipedia. The Kazakh Wikipedia we open very rarely, because there is not much in it. The one exception is when we need something from the history of Kazakhstan, because for that the other languages have little. Comparing three articles myself showed the same thing from the other side. On Astana the Russian article had 19 sections against 14 in Kazakh and was more recent, and the Kazakh one looked translated. On Mathematics the Russian article had 15 sections against 5, but the Kazakh one read as if it had been written in Kazakh rather than translated. On Gennady Golovkin the Russian article had 16 sections against 8, and the Kazakh article was again original rather than translated, with more attention to his state awards where the Russian one gave more space to international ones.

The following is the author's reaction after reading the results (Telegram, 25 September 2026), lightly edited and translated.

What surprised me most is that since 2013 not a single article has been created by a bot. I also did not expect that the rise in the human share of editing is not explained by more people arriving, but by bots leaving. And the most interesting thing is that the number of editors has actually fallen compared with 2011.

6. Limitations

The editor classes are properties of accounts. An unflagged account running a script is counted as human until the burst rule catches it, and a flagged bot performing hand-guided work is counted as a bot throughout. The burst rule narrows the first error and does nothing about the second.

The burst excess is an upper bound on unflagged automation, not an estimate of it. A genuine coordinated human effort — a school assignment, an edit-a-thon, a translation drive — produces the same shape in creation counts, and the behavioural corroborators reduce but do not eliminate that ambiguity: a hurried human stub is also short and also rarely revisited. The honest reading of Kazakh is a range, 27.2 to 48.3 per cent human creation, with our evidence favouring the lower half.

Steady automation is invisible to the rule by construction. An account creating two hundred pages a month for a decade never exceeds its own trailing median and is never flagged. The Uzbek and Russian placebo results (45 and 47 per cent) indicate that some flagged bot work in those editions has exactly this shape, so our corrections there are conservative.

The window begins in January 2010 and covers 70 to 99 per cent of each edition's live stock. The Tatar figure of 70 per cent is the weakest, so Tatar's position in Table 5 is the least secure in the set.

We measure creation and edits, not quality. A human-written page can be worse than a well-generated one, and nothing here should be read as a claim about the content of any article. The quality question would require article-level assessment, which this design does not attempt.

Finally, we do not identify any account, campaign or organisation. The attribution of the 2011 Kazakh bursts to the documented WikiBilim campaign rests on the coincidence of dates with a published record [12], not on account-level evidence, and should be treated as consistent with the data rather than established by it.

7. Conclusion

Of the content pages the Kazakh Wikipedia gained between 2010 and 2025, flagged bots created 50.4 per cent. Reading the burst signature as automation, the share created by a human being falls to 27.2 per cent, and that estimate holds across a tenfold range of the detection threshold. The same exercise puts the Kyrgyz figure at 16.6 per cent and the Tatar figure at 9.7 per cent, while the Azerbaijani edition, of near-identical size to the Kazakh one, was built almost entirely by hand.

The methodological result is the transferable one. The platform's own bot flag is a lower bound on automation and a poor one in exactly the editions where the question matters most, because mass creation also runs through ordinary accounts. The correction is cheap: monthly creation counts by editor class, a trailing-median rule, and two behavioural cross-checks that are already in the same API. It is falsifiable, it stays silent on the editions where nothing happened, and it recovers 87 to 97 per cent of known bot creation when used as a placebo.

For the specific argument this paper started from, the finding cuts both ways and we report both directions. Roughly half to three quarters of the Kazakh article counter is not human work, so the counter should not be quoted as a measure of Kazakh-language knowledge production without this decomposition [1]. But no bot has created a Kazakh article since 2012, the human share of editing has risen every few years since, and the recent flow is entirely human. What has not grown is the number of people doing it: the core Kazakh editing community is smaller today than during the 2011 campaign. The honest summary is that Kazakh Wikipedia's problem is no longer its bots. It is that roughly three dozen people a month are maintaining a quarter of a million articles, half of which somebody else's program wrote fifteen years ago.

Data availability and replication

All data are public and require no account. The pull script (`pull.py`), the analysis script (`analysis.py`) and the complete machine-readable results (`results.json`) accompany this manuscript. Running `pull.py` then `analysis.py` reproduces every number in every table from the Wikimedia Analytics API. The raw API payloads as retrieved on 23 September 2026 are included so that the tables can be reproduced without network access, and so that any later revision of the underlying data lake is detectable by comparison.

Declarations

Funding: none. Conflicts of interest: none. Ethics: the study uses only aggregate public counts; no personal data, no individual accounts and no human subjects are involved. Use of AI tools: the author used AI assistance for data retrieval code, statistical computation and language editing; the research question, the interpretation of results and the conclusions are the author's own, and every reported number traces to the saved scripts and payloads above.

References:

1. Zharkynbekova S., Ayupova G., Galiyeva B., Shakhputova Z., Zabrodskaja A. Digital linguistic sovereignty and Kazakh language policy in the digital environment // Frontiers in Psychology. 2026. DOI: 10.3389/fpsyg.2026.1858650.

2. Kornai A. Digital Language Death // PLoS ONE. 2013. Vol. 8, No. 10. e77056. DOI: 10.1371/journal.pone.0077056.

3. Steiner T. Bots vs. Wikipedians, Anons vs. Logged-Ins // Proceedings of the 23rd International Conference on World Wide Web Companion (WWW '14 Companion). 2014. arXiv:1402.0412.

4. Steiner T. Bots vs. Wikipedians, Anons vs. Logged-Ins (Redux): A Global Study of Edit Activity on Wikipedia and Wikidata // Proceedings of the International Symposium on Open Collaboration (OpenSym '14). 2014. DOI: 10.1145/2641580.2641613.

5. Tsvetkova M., García-Gavilanes R., Floridi L., Yasseri T. Even good bots fight: The case of Wikipedia // PLoS ONE. 2017. Vol. 12, No. 2. e0171774. DOI: 10.1371/journal.pone.0171774.

6. Halfaker A., Riedl J. Bots and Cyborgs: Wikipedia's Immune System // Computer. 2012. Vol. 45, No. 3. P. 79–82.

7. Geiger R. S. The Lives of Bots // Lovink G., Tkacz N. (eds.) Critical Point of View: A Wikipedia Reader. Amsterdam: Institute of Network Cultures, 2011.

8. Zheng L., Albano C. M., Vora N. M., Mai F., Nickerson J. V. The Roles Bots Play in Wikipedia // Proceedings of the ACM on Human-Computer Interaction. 2019. Vol. 3, CSCW, Article 215. DOI: 10.1145/3359317.

9. Reeves N., Simperl E. Machines in the Margins: A Systematic Review of Automated Content Generation for Wikipedia. 2025. arXiv:2509.22443.

10. Wikimedia Foundation. Analytics Query Service (Wikistats 2) REST API. URL: https://wikimedia.org/api/rest_v1/ (accessed 23.09.2026).

11. Wikipedia's largest non-English version was created by a bot. Generative AI poses new problems // ABC News. 15.04.2025. URL: https://www.abc.net.au/news/science/2025-04-15/wikipedia-cebuano-lsjbot-ai-article-generation-non-english/105123090 (accessed 23.09.2026).

12. Davis L. Class assignment inspired Wikipedian of the Year to grow Kazakh Wikipedia // Wikimedia Foundation, Diff. 05.07.2012. URL: https://diff.wikimedia.org/2012/07/05/class-assignment-inspired-wikipedian-of-the-year-to-grow-kazakh-wikipedia/ (accessed 23.09.2026).

UDC 699.841:711.4

Amangeldy Nurbolat

Grade 12 student
Nazarbayev Intellectual School of Physics and Mathematics, Shymkent
(Shymkent, Kazakhstan)

HOW MUCH OF KAZAKHSTAN'S URBAN BUILT STOCK PREDATES ITS SEISMIC CODES, AND WHERE DOES THAT STOCK SIT RELATIVE TO HAZARD?

Abstract: Kazakhstan introduced modern seismic design requirements in 1982, revised them in 2006 and replaced them with a Eurocode-harmonised code in 2017, yet no published study reports how much of the country's present urban built-up area was laid down before those rules existed, or how that older stock is distributed across seismic hazard. This study answers both questions for all 32 Kazakhstani urban centres using only open global data: the European Commission's GHS-AGE R2025A product, which records the epoch in which each 100 m cell first became built-up, masked by the GHS Urban Centre Database polygons, crossed with peak ground acceleration from the Global Earthquake Model hazard map and with the database's own macroseismic intensity classes. Of 180,745 hectares of urban built-up area, 47.9 per cent dates from before 1980 and 58.5 per cent from before 1990. Nine centres sit at a peak ground acceleration of 0.15 g or above; they hold 37.4 per cent of the national urban built-up area and 29.2 per cent of the pre-1980 stock. Exposure, defined as pre-code area multiplied by hazard, is extremely concentrated: the top three centres account for 79.1 per cent of the national total and Almaty alone for 61 per cent, 5.2 times the next city. The relationship between hazard and the age of the stock runs in the opposite direction to intuition: across the 32 centres a higher hazard is associated with a *lower* pre-1980 share (Spearman −0.53; regression coefficient on log hazard −0.028, p = 0.002, controlling for city size), because the aseismic northern cities are the oldest by share while the high-hazard south has been rebuilt fastest. The practical implication is that a national seismic-retrofit priority cannot be set from either variable alone: the stock that is both old and shaken is a small, identifiable set of places, led by Almaty.

Keywords: seismic risk; built environment; construction era; Kazakhstan; remote sensing; GHS-AGE; urban exposure

Introduction

Shymkent is an old city, and it has many old apartment blocks that may not withstand a strong earthquake. The district where my grandparents live is literally called the Old Town. Many of its private houses are now being demolished and replaced by new residential complexes. Near my school there is a residential complex where the walls are visibly crumbling and there are large cracks, which worries me for the people who live there and for everyone who walks past. At the same time new blocks are going up in the city centre, such as Turan, Asar 2 and Nursat 2.

An earthquake does not damage a city uniformly. What it damages depends on how the buildings were designed, and that in turn depends on when they were built, because seismic design requirements change over time and apply only to what is built after them. Kazakhstan's

requirements have three dates. The Soviet code SNiP II-7-81, *Construction in seismic regions*, came into force on 1 January 1982. It was replaced for independent Kazakhstan by SNiP RK 2.03-30-2006, and that in turn by SP RK 2.03-30-2017, *Construction in seismic zones*, in force from 20 December 2017 and harmonised with Eurocode 8. Everything standing today that was built before 1982 was therefore designed to a weaker standard than everything built after it, and in many cases to no explicit seismic standard at all.

The hazard itself is unevenly distributed. The south-east of the country, along the Tian Shan front, carries peak ground accelerations above 0.4 g, comparable to the most seismically active parts of Central Asia, while the northern and western plains are effectively aseismic. Almaty, the largest city, sits in the highest-hazard zone; Astana, the capital, does not.

Both facts are known. What has not been published is the intersection: how much of the built-up area of each Kazakhstani city dates from before the codes, and whether the cities with the most pre-code construction are the cities that shake. That intersection is what determines where retrofit money would do most good, and it can now be estimated without a single field survey, because the European Commission's Joint Research Centre publishes a global satellite-derived product recording the epoch in which every 100 m cell first became built-up, and the Global Earthquake Model publishes a global hazard map on a comparable grid.

This study crosses those two products for all 32 Kazakhstani urban centres and asks three questions. What share of the present urban built stock predates the 1982 code? How is that pre-code stock distributed across hazard, and which cities concentrate it? And does the age of a city's stock relate to its hazard at all, once city size is accounted for?

Materials and Methods

Unit of analysis. The 32 Kazakhstani urban centres defined by the GHS Urban Centre Database R2024A (GHS-UCDB V1.2). The urban centre, rather than the administrative city boundary, is used because it is defined consistently worldwide from population and built-up density, so the comparison between cities does not depend on where an administrative line happens to fall. Each centre carries a population for 2025 and a macroseismic intensity class from the same database.

Construction era. GHS-AGE R2025A at 100 m resolution records, for every built-up cell, the five-year epoch in which it first became built-up, from before 1975 through 2015–2020. Each cell is one hectare. The product was masked by each urban-centre polygon and the cells tallied by epoch class, giving for every city the built-up area in each era. Four era bands are used, mapped onto Kazakhstan's code dates at the nearest class boundary: pre-code (before 1980, approximating the 1982 code), Soviet-code era (1980–2005), first Kazakh code (2005–2015, approximating 2006–2017) and Eurocode-based (2015–2020, approximating 2018 onward). The mapping is imperfect by up to two years at each boundary because the satellite epochs do not fall on the code years; a sensitivity analysis shifting every boundary one class later is reported.

Hazard. Peak ground acceleration with a 10 per cent probability of exceedance in 50 years, on rock, from the Global Earthquake Model global seismic hazard map v2023.1, averaged over each urban-centre polygon. The database's macroseismic intensity class is carried alongside as an ordinal check. Almaty is the only Kazakhstani centre that the database splits across two intensity classes, 7 and 8, with 1.09 million residents in the lower and 1.22 million in the higher.

Measures. For each city: total built-up area in hectares; the share of that area in each era band; and an exposure index defined as pre-code built-up area multiplied by hazard, $E = A_{pre1980} \times PGA$, in hectare-g. The index is deliberately simple and is not a loss estimate: it

is a ranking device that answers "where is there the most old building in the places that shake hardest".

Statistics. The association between hazard and the age of the stock is reported three ways: Pearson and Spearman correlations between PGA and the pre-1980 share, and an ordinary-least-squares regression of the pre-1980 share on log hazard and log population with heteroskedasticity-robust (HC1) standard errors. With n = 32 the regression is reported as a descriptive association, not as a causal estimate, and the rank correlation is given because it does not assume linearity.

Decision rules, fixed before the analysis was run. Exposure is called "concentrated" only if the top three centres hold at least half of the national total. The hazard-versus-age relationship is claimed only if the regression coefficient on log hazard and the Spearman correlation are both significant at 5 per cent and agree in sign; otherwise it is reported as no association.

What is not claimed. GHS-AGE records the surface age of built-up land, not the structural type, material or code compliance of any building; the paper therefore speaks of "built-up surface dating from era X", never of "buildings that violate code Y". The product records first appearance, so demolition and replacement on the same footprint are invisible and the pre-code share is, if anything, an overestimate of what physically stands from that era. The GEM hazard grid is coarse relative to a city, so within-city variation is captured only through the UCDB two-class split for Almaty. Nothing after the product's last observed imagery, about 2018, is asserted.

Reproducibility. The per-city class tallies, the analysis script (Python, standard library only) and the full numerical output are released with this paper.

Results

The national picture. The 32 urban centres hold 180,745 hectares of built-up area. Of that, 47.9 per cent dates from before 1980 and 58.5 per cent from before 1990. Just under half of urban Kazakhstan, by surface, therefore predates the first modern seismic code.

Where the hazard is. Nine of the 32 centres sit at a polygon-mean peak ground acceleration of 0.15 g or above. Together they hold 37.4 per cent of the national urban built-up area and 29.2 per cent of the national pre-1980 stock. The remaining 23 centres, including the capital, are at or below that threshold, and the 22 centres below 0.05 g are effectively aseismic.

Table 1 – The ten Kazakhstani urban centres with the highest seismic exposure of pre-code built stock, 2026. Exposure index = pre-1980 built-up area × peak ground acceleration.

Rank	Urban centre	Built-up, ha	Pre-1980 share, %	PGA, g	Exposure index, ha·g	Population 2025
1	Almaty	30,313	43.8	0.399	5,300	2,315,538
2	Shymkent	21,598	28.5	0.165	1,017	1,239,663
3	Taraz	8,675	56.2	0.106	517	234,775
4	Taldykorgan	3,609	43.2	0.223	347	146,374
5	Esik	1,408	40.8	0.437	251	53,719
6	Zharkent	1,683	58.5	0.250	246	54,229
7	Sayram	2,472	29.4	0.167	122	71,026
8	Aqsukent	2,954	23.9	0.163	115	65,489
9	Jetisay	1,303	49.8	0.171	111	61,507
10	Ust-Kamenogorsk	5,516	54.6	0.036	108	240,226

Exposure is concentrated, and it is concentrated in one city. The national exposure total is 8,643 hectare-g. The top three centres hold 79.1 per cent of it and the top five 86.0 per cent, so the pre-registered threshold for "concentrated" is passed comfortably. Almaty alone accounts for 61 per cent, an exposure index 5.2 times that of Shymkent, the next city. Almaty holds 13,283 hectares of pre-1980 built-up area, 15.3 per cent of the national pre-1980 stock, and it sits at the highest hazard of any large Kazakhstani city. No other city combines the two.

Age runs against hazard, not with it. Across the 32 centres, higher hazard is associated with a *lower* pre-1980 share. The Spearman correlation is −0.53 and the Pearson correlation −0.45; in the regression, the coefficient on log hazard is −0.028 ($p = 0.002$) while city size is irrelevant (log population, $p = 0.965$), with an R^2 of 0.28. Both pre-registered conditions are met and they agree in sign, so the association is reported as real. The mechanism is visible in the extremes: the nine high-hazard centres average 38.3 per cent pre-1980 stock, the 22 aseismic centres 54.8 per cent. The oldest cities by share are Kostanay (73.4 per cent), Oral (67.1), Kokshetau (67.0), Semey (65.4) and Atyrau (64.2), all of them in the aseismic north and west. The fastest-renewing large cities are Shymkent, where 38.7 per cent of the built-up area appeared after 2005, Aktau (36.7) and Astana (32.2).

Table 2 – The hazard-versus-age association under three definitions of the pre-code era. All specifications control for log population with HC1 robust standard errors, n = 32.

Specification	Coefficient on log PGA	p	Spearman	R^2
Pre-1980 share (baseline)	−0.0278	0.002	−0.53	0.28
Pre-1990 share	−0.0228	0.006	−0.49	0.24
Boundary shifted one class later (pre-1985)	−0.0234	0.003	−0.48	—

Sensitivity. Moving the era boundary one class later, to pre-1985, leaves the association intact (coefficient −0.023, $p = 0.003$; Spearman −0.48), as does using the pre-1990 share instead (−0.023, $p = 0.006$). The finding is therefore not an artefact of where the code year is mapped onto the satellite epochs.

Almaty split by intensity. Almaty is the single centre the database divides between intensity classes 7 and 8, with 1.22 million of its 2.32 million residents in the higher class. The city's overall pre-1980 share, 43.8 per cent, is above the average for high-hazard centres, and its absolute pre-code area exceeds that of the next nine high-hazard centres combined. Resolving the era composition of the two intensity bands separately requires the intensity polygons rather than the aggregate class counts and is the natural extension of this work.

Discussion

Half of urban Kazakhstan, by built surface, predates the country's first modern seismic code, and that fact alone says little about risk, because most of that old surface stands where the ground does not shake. The distribution matters more than the total, and the distribution turns out to be lopsided in two ways at once.

First, it is lopsided between cities. Three centres hold four-fifths of the national exposure and one city holds three-fifths of it. A national retrofit programme that treated all old building as equally urgent would spend most of its money in Kostanay, Oral and Semey, the cities with the oldest stock by share, where the hazard is close to zero. The exposure ranking points instead

at a short list led by Almaty, with Shymkent and Taraz behind it and several small southern towns, Esik and Zharkent among them, appearing far above their population weight because they are both old and close to the Tian Shan front.

Second, it is lopsided in a direction that is easy to get backwards. The high-hazard south is not the country's oldest built environment; it is among its newest, because Shymkent, Taldykorgan and Turkestan have been rebuilt and extended rapidly since 2005, while the aseismic north has grown slowly and kept what it had. The negative association between hazard and age is therefore not a comforting result. It means that the aggregate national statistic, "half the stock is pre-code", is a poor guide to anything, and that the cities which most need attention are the exceptions to it, above all Almaty, which combines the highest hazard among large cities with an absolute pre-code area unmatched anywhere in the country.

The limits of the method should be read alongside these results. The satellite product sees the first appearance of built-up surface, not buildings; a 1960s footprint rebuilt in 2015 still counts as pre-1980, which biases the pre-code share upward and makes this an upper bound rather than a point estimate. Hazard is taken as a single value per city from a grid coarser than a city, so within-city differences, which are exactly what a municipal engineer needs, are beyond this design except for the Almaty split. And code dates are mapped to five-year epochs, so the boundary is approximate by up to two years, though the sensitivity analysis shows the conclusion does not hinge on it.

The epoch assignment was not checked against dated buildings on the ground. Comparing the satellite epoch of named buildings with their known construction years is the most direct next test of the upper-bound bias described above, and is left for further work.

Conclusion

Across Kazakhstan's 32 urban centres, 47.9 per cent of the 180,745 hectares of built-up area predates the 1982 seismic code and 58.5 per cent predates 1990. Nine centres sit at a peak ground acceleration of 0.15 g or above and hold 29.2 per cent of that pre-code stock. Exposure, measured as pre-code area times hazard, is concentrated to an unusual degree: 79.1 per cent of the national total in three cities and 61 per cent in Almaty alone, 5.2 times the next city. Across cities, hazard and the age of the stock are negatively associated (Spearman −0.53; −0.028 on log hazard, $p = 0.002$), because the aseismic north is the oldest by share while the high-hazard south has renewed fastest, and this holds under every alternative definition of the pre-code era tested. The policy reading is narrow and, for that reason, usable: the question "where is Kazakhstan's seismically exposed old building" has a short answer with a named list at the top of it, and that list can be produced for any country from open global data in a day, without a field survey.

Acknowledgments

Built-up epoch data are published by the European Commission Joint Research Centre (GHS-AGE R2025A, GHS Urban Centre Database R2024A) and the hazard map by the Global Earthquake Model Foundation (Global Seismic Hazard Map v2023.1); both are openly licensed and available to any reader. The analysis script and the full numerical output are released with this paper. The author declares no conflict of interest and received no funding for this work.

Use of AI tools. The author used generative AI tools to help write the analysis code, run the computations and edit the English text. The research direction, the observations of Shymkent in the author's own account and the choice of question are the author's. Every number reported here is produced by the released analysis script from the published JRC and GEM data; none was generated by a language model.

References:

1. Pesaresi M., Politis P. GHS built-up surface and age of the built environment, multitemporal (1975–2030). European Commission, Joint Research Centre (JRC), 2025. https://doi.org/10.2905/9F06F36F-4B11-47EC-ABB0-4F8B7B1D72EA

2. Florczyk A. J., Corbane C., Ehrlich D. et al. GHS Urban Centre Database, multitemporal and multidimensional attributes. European Commission, Joint Research Centre (JRC), 2024. https://data.jrc.ec.europa.eu/dataset/ghs-ucdb

3. Pagani M., Garcia-Pelaez J., Gee R. et al. The 2023 global seismic hazard map of the Global Earthquake Model. GEM Foundation, Zenodo, 2023. https://doi.org/10.5281/zenodo.8409647

4. SNiP II-7-81. Construction in seismic regions. USSR State Committee for Construction, in force from 1 January 1982.

5. SP RK 2.03-30-2017. Construction in seismic zones. Committee for Construction and Housing and Communal Services, Republic of Kazakhstan, in force from 20 December 2017.

6. Silva V., Amo-Oduro D., Calderon A. et al. Development of a global seismic risk model. Earthquake Spectra, 2020, vol. 36, no. 1_suppl, pp. 372–394. https://doi.org/10.1177/8755293019899953

7. Albiyev A. Decoupling gas flaring from oil production in Kazakhstan: a field-level satellite analysis, 2012–2025. Central Asian Scientific Journal, 2026, no. 2(30), vol. 3, p. 133.

Central Asian Scientific Journal

UDC 711.4

Torekhan Sabrina
11th-grade student
Scientific supervisor: Nurbakova Gulia Serikmukhametovna
Physics teacher
Pifagor High School
(Almaty, Kazakhstan)

WHERE ALMATY WALKS ITS OWN WAY: INFORMAL SHORTCUTS IN OPENSTREETMAP AND THE DETOURS THEY AVOID

Abstract: Desire paths are the worn tracks people make when the planned walking network does not follow the route they want to take. They have been mapped at city scale in the United States, but, to our knowledge, not in Central Asia. Using the open OpenStreetMap database for Almaty (snapshot of 28 September 2026), we tested a simple least-effort prediction: informal shortcuts should appear where the walking network forces a long detour between two points. We first planned to compare city parks, but the map is too sparse for that: only 7 of 127 mapped parks contain any way tagged with an unpaved surface. The test was therefore run city-wide. Of 576 unpaved pedestrian ways, 78 are short enough to be shortcuts (600 m or less), touch the mapped walking network (footways and walkable streets) at both ends, and have a route between their ends. Walking between the two ends of these shortcuts along the network takes a median of 2.16 times the straight-line distance, compared with 1.38 for randomly chosen pairs of network points at the same distance within 1 km of each shortcut (Mann-Whitney $z = 4.65$ to 5.31 across 20 random baseline draws; matched sign test 56 of 78 above their own baseline, $p < 0.001$). The median shortcut saves 108 m of walking, and 76 of 78 save distance. The result holds when the length cut-off is changed to 300 m or 1,000 m and when neighbouring shortcuts are grouped. It is consistent with least-effort theory, but it rests on the shortcuts volunteers have mapped, and the most important finding about the data is that most of Almaty's informal paths are not yet on the map.

Keywords: desire paths; pedestrian behaviour; OpenStreetMap; route directness; walkability; Almaty

Introduction. A desire path is a track worn into grass or earth by people who repeatedly take a route that the designed network does not provide. Such paths are cheap-to-observe evidence of where a public space disagrees with the way people use it. In Detroit, Foster and Newell mapped 5,680 informal footpaths from aerial imagery, with a combined length of about 157 miles, and linked them to vacant land and to trips toward everyday destinations such as schools and bus lines [1]. A study of a closed desire path found that it was used about four times more several months after closure than immediately after, suggesting that closing a shortcut does not remove the demand for it [2]. Agent-based models have also been developed to predict where desire paths will form in small neighbourhoods [3].

The common explanation is least effort: pedestrians minimise walking distance, and repeated shortcutting across open ground wears a visible track. Transport research measures how indirect a network is with circuity, the ratio of the network distance between two points to the straight-line distance [4, 5]. If least effort is right, informal paths should appear where circuity is high, that is, where the designed network makes walkers go a long way around.

This study tests that prediction in Almaty using data that already exist online: do the ends of Almaty's mapped informal shortcuts sit where the walking network makes people detour more than it usually does? It also measures how well OpenStreetMap covers informal paths in the city, which matters for any later work on this subject. OpenStreetMap is known to be close to complete for roads in many countries [6], but its coverage of small informal paths has not been assessed for Almaty.

Materials and Methods. Data. All pedestrian ways (OpenStreetMap tags highway = footway, path or pedestrian), all walkable streets (residential, living street, service, unclassified, tertiary, secondary and primary roads, their links, steps, tracks and corridors) and all mapped parks (leisure = park, including multipolygon relations) inside the Almaty administrative boundary were downloaded through the public Overpass API [7]. All three downloads have the data timestamp 2026-09-28T01:41Z. OpenStreetMap data are openly licensed under the ODbL [8].

Informal ways and the walking network. A way was classified as informal if its surface was tagged as ground, dirt, compacted, unpaved, grass, earth, mud or sand, or if it carried informal = yes. The walking network was defined as all footways, pedestrian ways and walkable streets that are not informal and are not marked as closed to pedestrians or private, plus paths with a recorded surface. Streets are included because in Almaty most sidewalks are not drawn as separate ways but are part of the street itself. Paths without any surface tag (764 ways) were left out of both groups, because they cannot be classified as either designed or informal. The resulting network has 97% of its points in one connected piece.

Which shortcuts can be tested. Network lines were divided into points every 10 m, so that the end of a shortcut can be attached to the nearest point of a line and not only to a junction. An informal way was tested if it is at most 600 m long (longer ways are trails rather than shortcuts), if each end lies within 25 m of the network, if its two ends attach to network points at least 10 m apart, and if a network route between them exists within 3 km.

Detour measure. For each tested shortcut, the shortest route along the network between its two attachment points was computed with Dijkstra's algorithm, and divided by the straight-line distance between the same two points. This ratio is the circuity of the network between the ends of the shortcut: 1.0 means the network is perfectly direct, higher values mean a longer detour. Walking distance saved is the network route, plus the short connections from each end to the network, minus the length of the shortcut itself.

Comparison. For each shortcut, 10 pairs of network points were drawn at random from the area within 1 km of the shortcut, with the two points of a pair at the same straight-line distance as the shortcut's ends (within 15%). Their circuity was computed in the same way, from network point to network point, so that shortcuts and comparison pairs are measured identically. Circuity of shortcut ends was compared with the comparison pairs using a two-sided Mann-Whitney test, repeated for 20 different random draws of comparison pairs. Because each shortcut has its own comparison pairs, a matched sign test was also used: for each shortcut, whether its circuity is above or below the median of its own 10 comparison pairs.

How the design changed. The first plan, written in September 2026, was to compare desire paths inside parks, starting with Roshcha Bauma, and to test the angle of path corners and nearby attractors such as entrances. When the data were downloaded, too few parks contained any mapped informal path to support that plan (Results). A first city-wide version of the test used footways only and was then corrected to include walkable streets, because without streets the

network was broken into thousands of disconnected pieces. The results reported here are from the corrected network.

Results. Coverage of the map. The city layer contains 21,434 pedestrian ways: 19,889 footways, 1,337 paths and 208 pedestrian streets. A surface is recorded for only 4,581 of them. Of the 127 mapped parks, 73 contain any pedestrian way, 1,650 ways in total, and 1,228 of these have no surface recorded. Only 12 parks contain any way tagged as a path, and only 7 contain any way tagged with an unpaved surface; 48 of those 74 unpaved ways are in Roshcha Bauma. A park-by-park comparison was therefore not possible with the current map.

Which shortcuts could be tested. Of 576 informal ways in the city, 196 are longer than 600 m, 278 have at least one end more than 25 m from the walking network, 10 have both ends attaching to practically the same point, and 14 have no network route between their ends within 3 km. The remaining 78 shortcuts were tested. Their median length is 88 m.

The detours they avoid. Walking between the ends of these shortcuts along the network takes a median of 225 m. The median circuity is 2.16 (interquartile range 1.31 to 3.95), compared with 1.38 (1.07 to 2.22) for the 780 comparison pairs of the first draw (Figure 1). Across 20 random draws of comparison pairs, the Mann-Whitney z ranges from 4.65 to 5.31, all $p < 0.001$. In the matched test, 56 of the 78 shortcuts have a higher circuity than the median of their own comparison pairs and 22 a lower one (sign test $p < 0.001$). Of the shortcuts, 68% sit where the network route is at least 1.5 times the straight line, compared with 40% of comparison pairs.

Distance saved. Seventy-six of the 78 shortcuts are shorter than the network alternative (Figure 2). The median saving is 108 m (interquartile range 29 to 316 m). Three shortcuts save more than one kilometre; each connects two points that the mapped network joins only by a long way around, which may reflect a real gap in the network or a missing link on the map.

Sensitivity. With a length cut-off of 300 m, 71 shortcuts are tested and the median circuity is 2.18 against 1.36 for comparison pairs (z from 4.48 to 5.01 across draws); with 1,000 m, 80 shortcuts, 2.16 against 1.34 (z from 4.78 to 4.91). Some shortcuts lie next to each other and may describe one crossing. Keeping only one shortcut in each group within 100 m leaves 61, with a median circuity of 2.18 and $z = 4.52$.

Discussion. The ends of Almaty's mapped informal shortcuts are located where the walking network is unusually indirect, and almost every shortcut makes the walk shorter. This is what least-effort theory predicts, and it agrees with the Detroit study, where informal paths were linked to places the planned network does not serve [1]: people cut their own paths where the designed route asks them to walk too far. The difference is not small. At the ends of a typical shortcut, the network route is more than twice the straight line, while for comparable points nearby it is about one and a third.

The more important finding, however, is about the map itself. Informal paths are one of the clearest signs of how a city is really used, yet in Almaty they are almost absent from the open map. Only 7 of 127 mapped parks show any unpaved way, and three quarters of pedestrian ways in parks have no surface recorded at all. The shortcuts that could be tested are only the ones that volunteers happened to draw and tag. Any analysis of desire paths in Almaty, by researchers or by the city, will need more complete mapping or another source, such as open satellite imagery.

Limitations. The tested sample (78 shortcuts) is not a random sample of Almaty's desire paths: it contains only paths that volunteers mapped and tagged, and it may over-represent the most visible ones. Whether a way is informal is decided by its surface tag, not by direct

observation. The walking network treats every street as walkable, although some wide roads are hard to cross, and it may still have missing links, which is one explanation for the few very large savings. The data show where shortcuts exist, not how many people use them. The analysis tests only one predictor, the detour; other factors that may matter, such as the angle of a corner or the pull of an entrance or bus stop, were not tested. The findings describe association and do not prove why each path was made.

Conclusion. In Almaty's open map, informal pedestrian shortcuts appear where the walking network makes people detour most: the network route between their ends is a median 2.16 times the straight line, against 1.38 for comparable nearby points, and the typical shortcut saves about 110 m. The result holds across random draws, length cut-offs and grouping of neighbouring shortcuts. The same data show that most of the city's desire paths are not yet mapped, which is the first obstacle for any study, or any planning decision, that wants to use them.

Data availability. All data are publicly available from OpenStreetMap through the Overpass API; the queries and the analysis code are available from the author on request.

Conflict of interest. The author declares no conflict of interest.

References:

1. Foster A., Newell J.P. Detroit's lines of desire: Footpaths and vacant land in the Motor City // Landscape and Urban Planning. 2019. Vol. 189. P. 260-273. DOI: 10.1016/j.landurbplan.2019.04.009.

2. Saxena N., Rashidi T.H., Babana J., Cheung C. Pedestrian Characteristics That Favor Desire Lines Despite Closure // Journal of Urban Planning and Development. 2020. Vol. 146, No. 2. DOI: 10.1061/(ASCE)UP.1943-5444.0000577.

3. Bossowski J., Szandała T., Mazurkiewicz J. Predicting Desire Paths: Agent-Based Simulation for Neighbourhood Route Planning // Computers, Environment and Urban Systems. 2025. URL: https://www.sciencedirect.com/science/article/pii/S0198971525000043 (accessed: 28.09.2026).

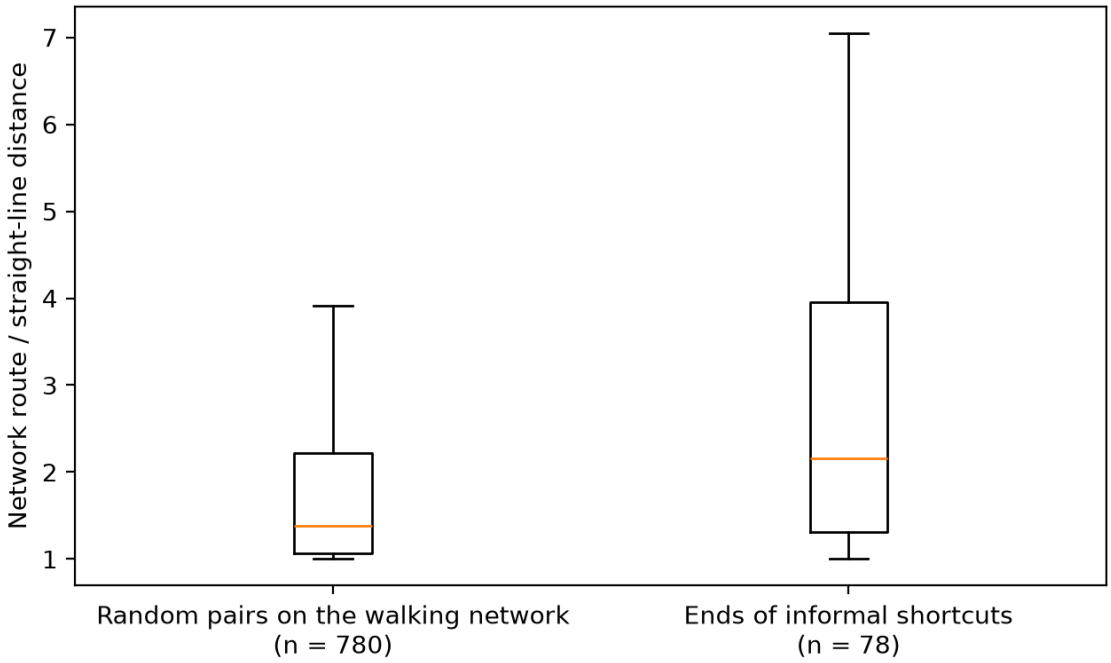
4. Randall T.A., Baetz B.W. Evaluating pedestrian connectivity for suburban sustainability // Journal of Urban Planning and Development. 2001. Vol. 127, No. 1. P. 1-15.

5. Levinson D., El-Geneidy A. The minimum circuity frontier and the journey to work // Regional Science and Urban Economics. 2009. Vol. 39, No. 6. P. 732-738.

6. Barrington-Leigh C., Millard-Ball A. The world's user-generated road map is more than 80% complete // PLoS ONE. 2017. Vol. 12, No. 8. Art. e0180698. DOI: 10.1371/journal.pone.0180698.

7. Overpass API. URL: https://overpass-api.de (accessed: 28.09.2026).

8. OpenStreetMap contributors. OpenStreetMap data, Open Database License (ODbL). URL: https://www.openstreetmap.org/copyright (accessed: 28.09.2026).



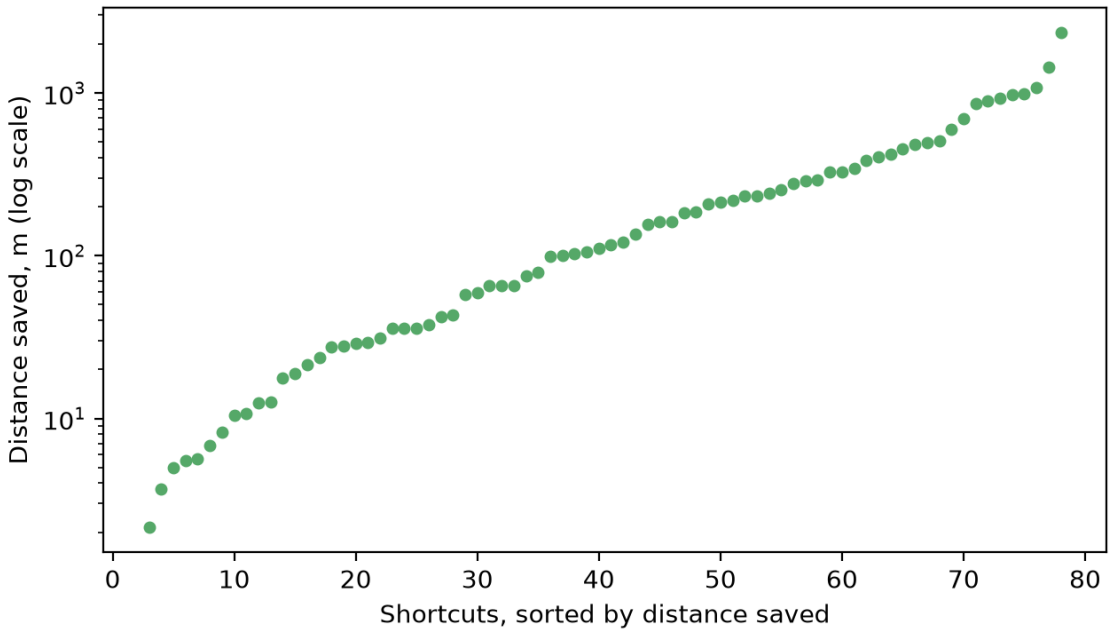


Figure captions:

Figure 1. Circuity (network route divided by straight-line distance) between the ends of 78 informal shortcuts and for 780 random comparison pairs of network points at matched distances within 1 km. Boxes show the interquartile range; values above the whiskers are not drawn.

Figure 2. Walking distance saved by each informal shortcut, sorted, on a logarithmic scale. Two shortcuts with no saving are not shown.

ӘЛЕУМЕТТІК ЖӘНЕ ГУМАНИТАРЛЫҚ ҒЫЛЫМДАР - СОЦИАЛЬНО - ГУМАНИТАРНЫЕ НАУКИ - SOCIAL SCIENCES AND HUMANITIES

ӘОЖ 001:004.8:1

Мұқатаева Айзат Амантайқызы

ф.ғ.к., педагогикалық академияның корр.-мүшесі
Тарих және өнер кафедрасы
«Шәкәрім университеті» коммерциялық емес акционерлік қоғамы
(Семей қ, Казахстан)

Стамбекова Жанат Байғазықызы,

Слямханова Гүлім Болатқызы,

Қызырбек Аяулым Ерболқызы,

1-курс магистранттар
Химия-биология кафедрасы
«Шәкәрім университеті» коммерциялық емес акционерлік қоғамы
(Семей қ, Казахстан)

ЖАСАНДЫ ИНТЕЛЛЕКТ ДӘУІРІНДЕГІ ҒЫЛЫМИ ТАНЫМ: ПОППЕР, КУН ЖӘНЕ ЛАКАТОС ИДЕЯЛАРЫ ТҰРҒЫСЫНАН ТАЛДАУ

Аңдатпа: Мақалада жасанды интеллекттің қазіргі ғылымның дамуына ықпалы ғылым тарихы мен философиясы тұрғысынан талданады. Зерттеудің негізгі мақсаты – жасанды интеллектті тек техникалық құрал ретінде емес, ғылыми танымның құрылымын, дәлелдеу тәсілдерін, гипотеза ұсыну логикасын және зерттеушінің жауапкершілігін өзгертетін эпистемологиялық фактор ретінде қарастыру. К. Поппердің фальсификация қағидасы, Т. Кунның парадигмалар теориясы, И. Лакатостың ғылыми-зерттеу бағдарламалары тұжырымдамасы және Р. Мертонның ғылым нормалары жасанды интеллект қолданылатын зерттеу тәжірибесімен салыстырылады. AlphaFold жүйесі, автономды зертханалар және генеративті модельдер мысалында ЖИ-дің деректерді талдау, модель құру, гипотеза генерациялау және экспериментті жоспарлау мүмкіндіктері көрсетіледі. Сонымен қатар «қара жәшік» мәселесі, алгоритмдік қателік, түсіндірмелілік, авторлық, ғылыми жауапкершілік және «түсінудің иллюзиясы» сияқты философиялық проблемалар талқыланады. Қазақстандағы 2024-2029 жылдарға арналған ЖИ дамыту тұжырымдамасы және 2025 жылғы «Жасанды интеллект туралы» Заң аясында ғылыми егемендік, қазақ тіліндегі деректер мен адамға бағдарланған реттеу мәселелері қарастырылады. Нәтижесінде ЖИ ғылымдағы адам субъектісін алмастыратын автономды «ғалым» емес, ғылыми пайымдаудың шекарасын кеңейтетін, бірақ оның сенімділігі адамдық сын, түсіндіру және жауапкершілік арқылы қамтамасыз етілуі тиіс күрделі эпистемикалық инфрақұрылым деген қорытынды жасалады.

Түйін сөздер: жасанды интеллект, ғылым философиясы, эпистемология, ғылыми этика, AlphaFold, алгоритмдік жауапкершілік.

Ғылым тарихында жаңа аспаптардың пайда болуы ғылыми танымның мазмұнын ғана емес, оның философиялық өзін-өзі түсінуін де өзгертіп отырды. Телескоп астрономиядағы бақылаудың шекарасын кеңейтсе, микроскоп тірі табиғаттың бұрын көрінбеген деңгейлерін ашты, компьютер күрделі есептеулер мен модельдеуді ғылыми тәжірибенің күнделікті бөлігіне айналдырды. Жасанды интеллект осы желінің жалғасы болғанымен, оның ерекшелігі тек жаңа объектілерді «көруге» көмектесуінде емес. Қазіргі машиналық оқыту жүйелері деректер арасындағы жасырын заңдылықтарды анықтайды, болжам ұсынады, мәтіндік гипотезаларды құрастырады, эксперимент параметрлерін оңтайландырады және кейбір жағдайларда келесі зерттеу қадамын таңдауға қатысады [1] [2]. Сондықтан ЖИ-дің ғылымдағы орны туралы мәселе техникалық тиімділік шеңберінен шығып, «ғылыми білімді кім өндіреді?», «түсіндіру мен болжаудың арақатынасы қандай?», «алгоритм ұсынған нәтиженің ақиқаттығы қалай тексеріледі?» деген классикалық философиялық сұрақтарды жаңа формада қайта қояды. Бұл мәселенің өзектілігін ғылыми өндірістің сандық динамикасы да көрсетеді. Stanford Human-Centered AI Institute жариялаған 2025 жылғы AI Index есебіне сәйкес, жасанды интеллектке қатысты ғылыми жарияланымдар саны 2013 жылғы шамамен 102 мыңнан 2023 жылы 242 мыңнан астамға дейін өскен, ал компьютерлік ғылымдар ішіндегі ЖИ жарияланымдарының үлесі 21,6%-дан 41,8%-ға жеткен [3]. Мұндай өсім ЖИ-дің дербес зерттеу нысаны ғана емес, биологиядан материалтануға, медицинадан әлеуметтік ғылымдарға дейін таралған әмбебап ғылыми әдіснамаға айналып келе жатқанын аңғартады. 2024 жылғы химия бойынша Нобель сыйлығының бір бөлігі Демис Хассабис пен Джон Джамперге ақуыз құрылымын болжауға арналған AlphaFold жүйесі үшін берілуі ЖИ-дің ғылыми жаңалық жасаудағы институционалдық мәртебесінің өзгергенін айқын көрсетті [4].

Сонымен қатар ЖИ-дің ғылыми ортаға енуі «технологиялық оптимизмге» ғана негізделген біржақты көзқарасты көтермейді. Үлкен деректер мен нейрондық желілер жоғары болжамдық дәлдікке қол жеткізгенімен, модельдің шешімге қалай келгені әрдайым теориялық тұрғыдан түсіндірілмеуі мүмкін. Бұл жағдай ғылымдағы түсіндіру идеалы мен инженерлік тиімділік идеалының арасындағы қайшылықты күшейтеді. Мессери мен Крокетт ғылыми зерттеулердегі ЖИ жүйелері зерттеушілерде «түсінудің иллюзиясын» қалыптастыруы мүмкін екенін көрсетеді – құрал көп нәрсені есептеп, жіктеп немесе болжай алған сайын, адам сол нәтижені теориялық жағынан жеткілікті түсіндім деп қате ойлауы ықтимал [5]. Осыдан ғылыми білімнің көлемі артқанымен, оның мағыналық тереңдігі автоматты түрде артпайтыны байқалады.

Мақаланың мақсаты – жасанды интеллекттің ғылым дамуындағы рөлін ғылым философиясының негізгі категориялары арқылы талдап, оның ғылыми танымның әдіснамасы, эпистемологиясы, әлеуметтік нормалары мен этикасына әкелетін өзгерістерін жүйелеу.

Жасанды интеллекттің ғылымға ықпалын дұрыс бағалау үшін оны ғылым тарихындағы есептеу құралдарының эволюциясынан бөлек қарастыру жеткіліксіз. XVII ғасырдағы механикалық есептеу машиналары, XIX ғасырдағы статистикалық әдістер, XX ғасырдағы электронды компьютерлер зерттеушінің белгілі бір операцияларын сыртқы құрылғыға берді. Алайда бұл құрылғылардың логикасы негізінен алдын ала анықталған ережелерді орындаумен шектелді. Қазіргі машиналық оқыту жүйелерінің айырмашылығы – олар дайын ережелерді ғана орындамай, үлкен деректер массивтерінен

статистикалық құрылымдарды өздігінен үйренеді. Осы себепті ЖИ ғылымдағы «құрал» ұғымын кеңейтеді – ол өлшеу немесе есептеу қызметін атқарып қана қоймай, нысанды сипаттаудың жаңа репрезентацияларын, ықтимал гипотезаларды және эксперименттік іздеу кеңістігін қалыптастырады [6].

Бұл өзгерісті Әбу Насыр әл-Фарабидің ғылымдарды жіктеу дәстүрімен байланыстыра қарастыруға болады, өйткені ол ғылымды әр саланың өзіндік пәні, әдісі мен дәлелдеу формасы бар ұйымдасқан білім жүйесі ретінде түсіндірді [7]. Қазіргі ЖИ әртүрлі деректерді бір архитектура негізінде өңдеу арқылы ғылым салалары арасындағы әдіснамалық трансферді жеделдетеді. Оның ғылыми маңызы әсіресе ауқымды іздеу кеңістігінде байқалып, машиналық оқыту көптеген ықтимал нұсқалардың арасынан перспективалы бағыттарды анықтауға мүмкіндік береді. Осы тұрғыдан ЖИ ғылыми жаңалықты өздігінен ашатын жүйе емес, зерттеушінің іздеу кеңістігін ұйымдастыратын және маңызды деректерді анықтауға көмектесетін құрал болып табылады.

Карл Раймунд Поппер ғылымды теріске шығарылуы мүмкін болжамдар жүйесі ретінде қарастырып, ғылыми теорияның қатаң эмпирикалық тексеруге ашық болуын маңызды деп санады [8]. ЖИ модельдері жаңа заңдылықтар мен тексерілетін болжамдарды анықтауға мүмкіндік бергенімен, күрделі модельдердің ішкі механизмдерінің түсініксіздігі олардың теориялық негізін бағалауды қиындатуы мүмкін. Сондықтан ЖИ ұсынған болжамның эмпирикалық дәлдігімен қатар, оның қандай алғышарттарға негізделгені және қандай жағдайда теріске шығарылатыны да айқын болуы тиіс.

Томас Сэмюэл Кун ғылымның дамуында ғылыми қауымдастық қабылдаған парадигмалардың рөлін атап өтіп, олардың зерттеу сұрақтарын, әдістер мен деректерді түсіндіру тәсілдерін айқындайтынын көрсетті [9]. ЖИ парадигмалық құрылымға енгенде бейтарап құрал болып қалмай, деректерді, модель сапасының өлшемдерін және зерттеу басымдықтарын таңдауға ықпал етеді. Имре Лакатостың ғылыми-зерттеу бағдарламалары концепциясы бойынша теориялар «қатты өзек» пен өзгеретін қосалқы гипотезалар аясында дамиды, ал ЖИ негізіндегі зерттеулерде модель архитектурасы, параметрлері мен бағалау метрикалары осыған ұқсас өзгерістерге ұшырайды [10]. Сондықтан модельдің жоғары дәлдігі ғылыми түсіндіруді алмастырмай, теориялық негізбен және жаңа эмпирикалық мазмұнмен толықтырылуы тиіс.

ЖИ дәуіріндегі ғылыми әдіс Карл Раймунд Поппердің тексерілу және теріске шығарылу, Томас Сэмюэл Куннің парадигмалық алғышарттарды ашық көрсету және Имре Лакатостың прогрессивті зерттеу бағдарламасы қағидаларын сақтауды талап етеді. Осы тұрғыдан ЖИ ғылыми рационалдылыққа қарсы құбылыс емес, оның әдіснамалық талаптарын қайта қарастыруға ықпал ететін фактор ретінде көрінеді. ЖИ-дің ғылымдағы нақты рөлін бағалау үшін ғылыми қызметті мәселені қоюдан бастап деректерді жинау, модельдеу, гипотеза ұсыну, эксперимент жүргізу, нәтижені түсіндіру және жариялауға дейінгі біртұтас цикл ретінде қарастыру қажет, өйткені әр кезеңдегі ЖИ функцияларының мүмкіндіктері мен тәуекелдері әртүрлі.

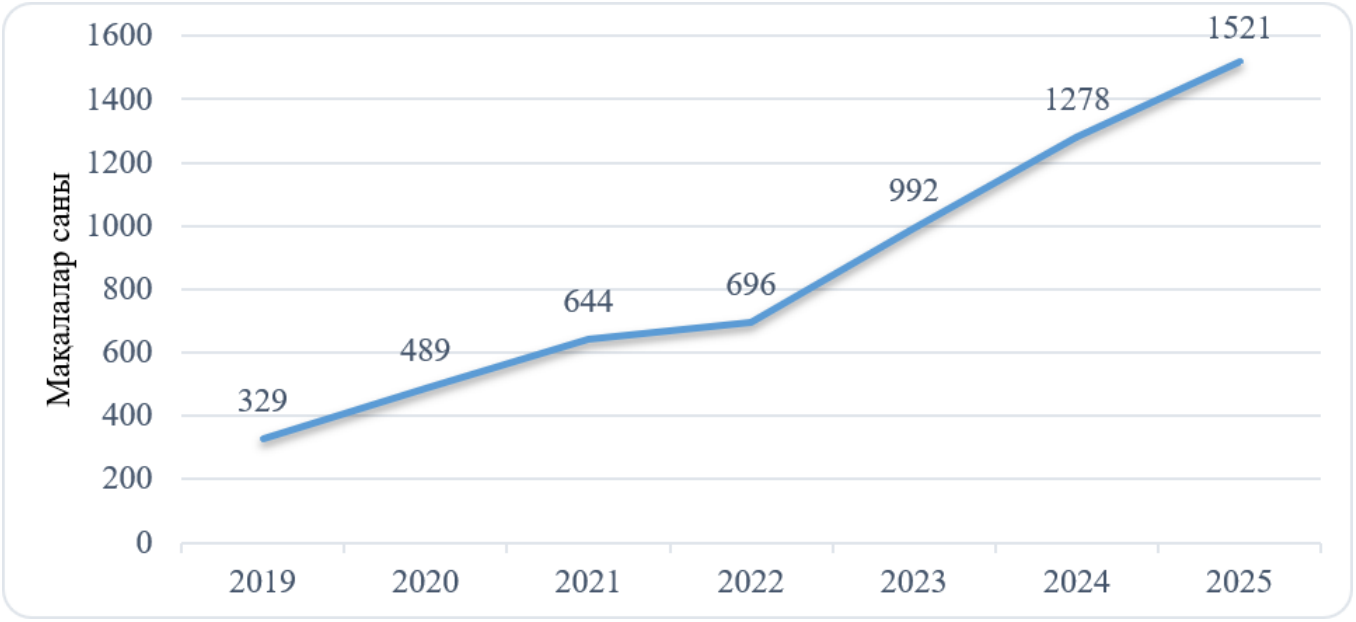
Кесте 1 – Ғылыми таным циклінің ЖИ ықпалымен трансформациясы

Ғылыми кезең	**ЖИ-дің функциясы**	**Эпистемикалық мүмкіндік**	**Негізгі философиялық тәуекел**
Мәселені қою	Әдебиетті семантикалық іздеу, зерттеу олқылықтарын белгілеу	Үлкен әдеби массивті тез байланыстыру	Маңызды мәселені «танымал» тақырыппен шатастыру
Дерекпен жұмыс	Тазалау, жіктеу, аномалия табу, белгілеу	Адам байқамайтын үлгілерді анықтау	Дерек сапасының жасырын қателіктерін көбейту
Модельдеу	Күрделі тәуелділіктерді үйрену, симуляция	Көп факторлы процестерді болжау	«Қара жәшік» және себептілік пен корреляцияны шатастыру
Гипотеза құру	Баламалы түсіндірмелер мен болжамдар ұсыну	Іздеу кеңістігін кеңейту	Статистикалық ықтималдықты теориялық негіз деп қабылдау
Эксперимент	Параметрлерді оңтайландыру, белсенді оқыту	Уақыт пен ресурсты үнемдеу	Оңтайландыру критерийінің ғылыми мақсатты алмастыруы
Интерпретация	Нәтижелерді салыстыру, үлгі табу	Күрделі деректерді құрылымдау	Түсінудің иллюзиясы және автоматтандырылған бедел
Коммуникация	Мәтінді жүйелеу, аудару, визуализация	Ғылыми ақпаратқа қолжетімділікті арттыру	Авторлық, дәйексөз және жалған сілтеме мәселелері
Ескерту – ғылыми танымдағы ЖИ функциялары бойынша қарастырылған еңбектердің идеялары негізінде құрастырылды			

1-кестеден көрінгендей, ЖИ-дің басты артықшылығы ғылыми циклдің жекелеген буындарын «автоматтандыруда» ғана емес, олардың арасындағы уақыттық байланысты қысқартуда. Бұрын әдебиетке шолу, дерек дайындау және экспериментті жоспарлау ұзақ уақыт аралығында кезекпен орындалса, қазіргі жүйелер бұл кезеңдердің бір бөлігін параллель жүргізуге мүмкіндік береді. Өздігінен басқарылатын зертханалар белсенді оқыту алгоритмдерін роботтандырылған экспериментпен біріктіріп, алдыңғы нәтиже негізінде келесі тәжірибені автоматты түрде таңдауға ұмтылады. Бұл ғылыми циклдің итерациялық табиғатын күшейтеді– «болжам – тәжірибе – түзету» тізбегі сағаттар немесе күндер ішінде бірнеше рет қайталануы мүмкін.

AlphaFold мысалы ЖИ-дің ғылыми танымдағы рөлін одан да айқын көрсетеді. Ақуыздың аминқышқылдық реттілігі бойынша оның үшөлшемді құрылымын болжау ондаған жыл бойы биологияның күрделі мәселелерінің бірі болды. AlphaFold2 жүйесі көптеген ақуыздардың құрылымын эксперименттік әдістерге жақын дәлдікпен болжауға мүмкіндік беріп, құрылымдық биологияда жаңа зерттеу мүмкіндіктерін ашты [11]. Кейін AlphaFold3 ақуыздармен қатар нуклеин қышқылдары, шағын молекулалар, иондар және модификацияланған қалдықтар қатысатын кешендердің құрылымын модельдеуге бағытталды [12]. 2024 жылғы Нобель сыйлығы бұл жетістіктің ғылыми маңызын

институционалдық түрде бекітті. Алайда AlphaFold мысалы «болжау = түсіндіру» теңдеуінің дұрыс еместігін де көрсетеді. Құрылымды жоғары дәлдікпен болжау ақуыздың барлық динамикалық, химиялық немесе жасушалық контекстін толық түсіндіреді деген сөз емес. Nature Methods журналындағы салыстырмалы талдау AlphaFold болжамдарының аса құнды ғылыми гипотеза екенін, бірақ эксперименттік құрылым анықтауды толық алмастырмайтынын көрсетеді [13]. Философиялық тұрғыдан ЖИ ғылымдағы теориялық және эмпирикалық процедураларды алмастырмайды, керісінше олардың өзара байланысын өзгертеді. ЖИ ғылыми іздеуді бағыттап, болжамдар қалыптастыру мүмкіндігін кеңейтеді, алайда оның тиімділігі деректердің сапасына, модель архитектурасына және зерттеу сұрағының дұрыс қойылуына тәуелді. Сондықтан ғылыми нәтижелердің сенімділігін бағалау, оларды түсіндіру және модель қателіктерін анықтау зерттеушінің маңызды міндеті болып қала береді. AI Index деректері ЖИ тақырыбындағы зерттеулер үлесінің айтарлықтай өскенін көрсетіп, оның жеке зерттеу бағытымен қатар әртүрлі ғылым салаларына ықпал ететін маңызды технологияға айналғанын аңғартады.



Сурет 1 – ЖИ тақырыбындағы жарияланымдардың компьютерлік ғылымдардағы үлесінің өсуі, 2019-2025 жж., %

Дереккөз: Stanford AI Index Report 2026. Деректер алты жетекші ЖИ конференциясында (AAAI, AIES, FAccT, ICML, ICLR, NeurIPS) қабылданған RAI бағытындағы мақалаларды қамтиды.

Ғылым философиясында білімнің сенімділігі тек нәтиженің дәлдігімен өлшенбейді. Ғалым неге белгілі бір қорытындыға келгенін, қандай алғышарттарды пайдаланғанын және оның қандай жағдайда өзгеруі мүмкін екенін көрсетуі тиіс. Көптеген терең нейрондық желілердің күрделі ішкі құрылымы осы талаппен қайшылыққа түседі. Модель дұрыс нәтиже беруі мүмкін, бірақ оның шешімі адамға түсінікті теориялық ереже түрінде ұсынылмауы ықтимал. «Қара жәшік» мәселесінің мәні алгоритмнің күрделі болуының өзінде емес, мәселе ғылыми дәлелдің қоғамдық тексерілуіне қажет себептік және концептуалдық байланыстың әлсіреуінде. Деректер ғылымының эпистемологиялық негіздерін талдаған Десаи, Уотсон, Таддео және

Флориди деректерге негізделген зерттеулер дәстүрлі теориялық ғылымның баламасы емес, өзіндік әдіснамалық алғышарттары бар таным режимі екенін көрсетеді [14]. Дерек көп болған сайын теорияның қажеті азаяды деген тезис тым қарапайым. Үлкен дерек те қандай айнымалылар өлшенгеніне, қандай категориялар алдын ала белгіленгеніне және қандай ақпарат жоғалғанына тәуелді. Алгоритм заңдылық тапқан кезде ол заңдылықтың табиғи, әлеуметтік немесе техникалық себептерін өздігінен ажырата алмайды. Сондықтан деректерден табылған құрылым ғылыми мағынаға тек теориялық контекст арқылы ие болады.

Мессери мен Крокетт сипаттаған «түсінудің иллюзиясы» осы жерде ерекше маңызға ие. Ғылыми қауымдастық интеллектуалдық міндеттерді автоматтандырған кезде зерттеуші белгілі бір құбылысты бұрынғыдан жақсы түсінетіндей сезінуі мүмкін. Бірақ модельдің қысқаша жауабы немесе әдемі визуализациясы мәселенің себептік механизмі түсіндірілді дегенді білдірмейді. Авторлар мұндай құралдар ғылыми зерттеудің өнімділігін арттырғанымен, зерттеушілердің бірдей типтегі әдістерге тәуелді болуын, альтернативті түсіндірулердің азаюын және эпистемикалық монокультура қалыптасуын да күшейте алатынын ескертеді. Яғни ғылым жылдамырақ жұмыс істеп, бірақ бір бағытта ғана ойлайтын жүйеге айналуы ықтимал.

Лючано Флориди ұсынған «сенімділік пен қамту ауқымы арасындағы айырбас» туралы тұжырым да осы мәселені тереңдетеді [15]. Символдық жүйелер тар әрі формалданған кеңістікте жоғары дәлелдік сенімділік бере алады, ал генеративті ЖИ өте кең тақырыптарды қамтығанымен, әр жауаптың формалды кепілдігі төмен. Ғылым үшін бұл айырбас практикалық мағынаға ие – әмбебап модельді көптеген салада қолдануға болады, бірақ оның әрбір нақты тұжырымы үшін сыртқы тексеру қажет. Осы себепті ғылыми ортада «модель айтты» деген аргумент дереккөз немесе дәлел ретінде қабылданбауы тиіс, модель тек тексерілуі тиіс аралық интеллектуалдық өнім ұсынады.

ЖИ-дің қателігі де дәстүрлі аспап қатесінен өзгеше. Өлшеу құралының қатесі көбіне статистикалық сипатта бағаланса, генеративті жүйе мазмұндық тұрғыдан сенімді көрінетін, бірақ жалған ақпарат құрастыра алады. Бұл қасиет ғылыми мәтіндегі дәйексөздер мен әдебиеттер тізімінде ерекше қауіпті. Жалған DOI, ойдан шығарылған автор немесе жоқ журнал ғылыми верификация процесін бұзады. Сондықтан ЖИ қолданылатын ғылыми жұмыста бастапқы дереккөзді тікелей тексеру, библиографиялық жазбаны дерекқордан растау және маңызды сандық деректерді бірнеше тәуелсіз көзбен салыстыру жаңа ғылыми сауаттылықтың міндетті бөлігіне айналады. Осыдан түсіндірмелілікке қатысты бір маңызды айырмашылық туындайды. Әрбір модельді толық «ішінен ашу» мүмкін болмауы мүмкін, бірақ ғылыми контексте оның қолданылу шекарасы, дерек шығу тегі, қате түрлері, бағалау метрикалары және шешімнің тұрақтылығы түсіндірілетін болуы керек. Яғни түсіндірмелілік тек техникалық ХАІ әдістерінің мәселесі емес, ол зерттеу дизайнының ашықтығын, деректердің құжатталуын және нәтижені қайта өндіру мүмкіндігін қамтитын кең ғылыми норма.

ЖИ туралы философиялық пікірталаста ең тартымды сұрақтардың бірі – машинаны таным субъектісі деп атауға бола ма деген мәселе. Джон Серлдің «қытай бөлмесі» аргументі синтаксистік ережелерді шебер орындау семантикалық түсінумен тең емес екенін көрсетуге бағытталған [16]. Қазіргі үлкен тілдік модельдерге қатысты бұл аргумент қайта өзектенді – жүйе ғылыми стильде дәлел құрып, ұғымдарды байланыстыра алады, бірақ оның мәтіндік өнімінен саналы түсіну бар деген қорытынды автоматты түрде шықпайды. Ғылым философиясы үшін маңыздысы – машинаның «санасы» туралы

алыпсатарлықтан гөрі, ғылыми қауымдастықта дәлелге жауап беретін субъект кім екені болып табылады.

Ғылыми авторлық тек мәтін өндіру қабілетіне негізделмейді. Автор зерттеу сұрағын таңдауға, деректің заңды және этикалық алынуына, әдістің дұрыстығына, нәтижені интерпретациялауға және қате анықталған жағдайда түзетуге жауап береді. ЖИ жүйесі бұл міндеттердің бір бөлігін техникалық түрде орындай алады, бірақ моральдық немесе құқықтық жауапкершілікті өз мойнына алмайды. Сондықтан қазіргі жағдайда ЖИ-ді автор емес, қолданылған интеллектуалдық құрал немесе әдістемелік компонент ретінде сипаттау қисынды. Бұл ұстаным ғылыми мәтінде ЖИ қолдануды жасыру емес, оның нақты функциясын көрсету қажеттігін туғызады.

Роберт Мертон ғылымның этосын универсализм, ортақтық, мүддесіздік және ұйымдасқан скептицизм сияқты нормалармен сипаттаған [17]. ЖИ осы нормалардың әрқайсысына жаңа сынақ әкеледі. Универсализм тұрғысынан модельдің шешімі зерттеушінің мәртебесіне емес, дәлел сапасына сүйенуі тиіс, алайда алгоритмдік жүйе оқыту дерегіндегі әлеуметтік теңсіздіктерді қайталай алады. Ортақтық қағидасы ғылыми білім мен әдістің ашықтығын қолдайды, бірақ ірі коммерциялық модельдердің жабық коды немесе деректер жиыны қайта өндіруді шектейді. Мүддесіздік қағидасына модель иесінің коммерциялық мақсаттары әсер етуі мүмкін. Ұйымдасқан скептицизм ЖИ жауаптарын беделді «дайын жауап» ретінде қабылдамай, тұрақты тексеруді талап етеді.

UNESCO-ның 2021 жылғы ЖИ этикасы жөніндегі ұсынымы адам құқықтары мен қадір-қасиетін негізгі бағдар ретінде бекітіп, ашықтық, әділдік және адамдық бақылауды маңызды қағида ретінде көрсетеді [18]. Ғылыми зерттеуде бұл қағидалар бірнеше нақты талапқа айналады – қатысушылардың деректерін қорғау, алгоритмдік дискриминацияны бағалау, қоршаған ортаға түсетін есептеу жүктемесін ескеру, ЖИ шешіміне соңғы ғылыми жауапкершілікті адамда сақтау. Этикалық бақылау ғылыми прогресті баяулататын сыртқы шектеу емес, керісінше, сенімсіз немесе қоғамға зиянды нәтижелердің алдын алатын ғылыми сапа механизмі.

Егер ЖИ жүйесі зерттеу бағытын өзі ұсынса, эксперимент параметрін таңдаса және нәтижені интерпретациялауға көмектессе, адам рөлі қайда қалады деген сұрақ туындайды. Мұнда «human-in-the-loop» ұғымы формалды батырма басумен шектелмеуі керек. Адамның мағыналық бақылауы зерттеу мақсатын таңдауда, этикалық шектеулерді белгілеуде, күтпеген нәтижені тануда және теориялық түсіндіруде көрінеді. Ғылымдағы жауапкершілік – тек соңғы шешімді бекіту емес, бүкіл зерттеу логикасын сын тұрғысынан түсіну қабілеті. Осы философиялық мәселелерді бір жүйеге келтіру ЖИ туралы пікірталастың тек «пайдалы/зиян» оппозициясынан әлдеқайда күрделі екенін көрсетеді. Негізгі сұрақ – технологияны қолдану немесе қолданбауда емес, оны қандай эпистемикалық және этикалық тәртіппен ғылымның ішіне енгізуде.

Кесте 2 – ЖИ қолданудың негізгі философиялық мәселелері және ғылыми жауап қағидалары

Өлшем	Негізгі сұрақ	Қауіп	Ғылыми жауап қағидасы

Эпистемология	Алгоритм нәтижесі білім болып қашан саналады?	Дәлдік пен ақиқаттылықты теңестіру	Тәуелсіз тексеру және дереккөзге оралу
Әдіснама	Модель ғылыми әдісті қалай өзгертеді?	Метриканы ғылыми мақсаттың орнына қою	Алдын ала хаттама, валидация, қайта өндіру
Онтология	ЖИ таным субъектісі ме, әлде құрал ма?	Агенттік пен жауапкершілікті антропоморфизациялау	Функция мен субъект мәртебесін ажырату
Этика	Қате немесе зиян үшін кім жауап береді?	Жауапкершіліктің «еріп кетуі»	Адамдық бақылау және рөлдерді нақтылау
Ғылым социологиясы	Жабық модельдер ғылым нормаларына қалай әсер етеді?	Эпистемикалық монополия және тәуелділік	Ашық әдіс, аудит, инфрақұрылымдық әртараптандыру
Мәдени-лингвистикалық өлшем	Қай тіл мен мәдениет деректе көрінеді?	Төмен ресурсты тілдердің шеттетілуі	Ұлттық корпус, жергілікті модель, көптілді валидация
Ескерту – ғылыми танымдағы ЖИ функциялары бойынша қарастырылған еңбектердің идеялары негізінде құрастырылды			

Қазақстан үшін ЖИ-дің ғылымдағы рөлі жаһандық трендті қайталау мәселесімен шектелмейді. Мұнда ғылыми және технологиялық егемендік, қазақ тіліндегі деректердің сапасы, ұлттық зерттеу инфрақұрылымы және құқықтық жауапкершілік қатар қарастырылуы қажет. Қазақстан Республикасы Үкіметінің 2024 жылғы 24 шілдедегі №592 қаулысымен бекітілген Жасанды интеллектті дамытудың 2024–2029 жылдарға арналған тұжырымдамасында деректер, инфрақұрылым, адами капитал және ЖИ саласындағы ғылыми зерттеулер жеке бағыттар ретінде белгіленген [19]. Құжатта KazLLM сияқты қазақ тіліне бағытталған ауқымды тілдік модельдерді дамыту қажеттілігі де көрсетіледі. Бұл бастама техникалық жоба ғана емес, ғылыми білімді ұлттық тілде өндіру мен тарату мүмкіндігіне тікелей қатысты.

Қазақстан Республикасының 2025 жылғы 17 қарашадағы №230-VIII «Жасанды интеллект туралы» Заңы 2026 жылы күшіне еніп, ЖИ қолданудың құқықтық негізін қалыптастырды [20]. Заңдағы адам, оның құқықтары, бостандықтары және әл-ауқатының басымдығы туралы ұстаным ғылыми зерттеуге де тікелей қатысты. Әсіресе медицина, білім, психология және әлеуметтік зерттеулерде ЖИ жүйесі адамның тағдырына әсер ететін қорытындылар шығарса, модельдің техникалық тиімділігі құқықтық және этикалық негізден бөлек бағаланбауы керек. Ғылыми жаңалықтың қоғамдық құндылығы оны қалай алғанымызға және қандай тәуекелмен қолданғанымызға байланысты.

Қазақстандық ғылыми орта үшін маңызды мәселелердің бірі – шетелдік жабық ЖИ модельдеріне инфрақұрылымдық тәуелділік, өйткені сыртқы платформалардың өзгеруі зерттеу нәтижелерінің қайта өндірілуіне әсер етуі мүмкін. Осыған байланысты ғылыми егемендік маңызды деректерді сақтау, модель нәтижелерін аудиттеу, балама құралдарға көшу және ұлттық ғылыми басымдықтарға сәйкес жергілікті құзыреттілікті дамыту мүмкіндігін қамтуы тиіс. Қазақстандағы ЖИ саясаты адамға бағдарлану, эпистемикалық тәуелсіздік және тілдік-мәдени инклюзивтілік қағидаларына негізделіп, қазақ тілі мен жергілікті контекст модель сапасын бағалаудың маңызды өлшемдерінің бірі болуы қажет. Сонымен қатар ЖИ-дің ғылымға ықпалын тек зерттеу жылдамдығы мен

өнімділігінің артуымен бағалау жеткіліксіз, өйткені ғылыми прогресс нәтижелердің сенімділігіне, теориялық негізділігіне, қайта өндірілуіне және қоғамдық маңыздылығына тәуелді [21].

ЖИ ғылыми зерттеулердің болжамдық мүмкіндіктерін кеңейткенімен, фундаменталды ғылымда құбылыстардың механизмін теориялық тұрғыдан түсіндіру негізгі талап болып қала береді. Адам мен алгоритмнің өзара әрекеттесуінде ЖИ үлкен деректерден жасырын заңдылықтарды анықтаса, зерттеуші олардың контекстін, ғылыми мағынасын және құндылығын түсіндіреді. Сондықтан ЖИ-ді дербес «ғалым» ретінде емес, ғылыми іздеу мүмкіндіктерін кеңейтіп, зерттеу үдерісін өзгертетін эпистемикалық инфрақұрылым ретінде қарастырған орынды.

ЖИ деректерді өңдеу, модель құру, гипотеза ұсыну және экспериментті оңтайландыру арқылы ғылыми зерттеу циклін жеделдеткенімен, ғылыми білімнің құндылығы тек болжамдық дәлдікпен шектелмейді. «Қара жәшік» модельдері, жалған генерация, деректік бұрмалану және эпистемикалық монокультура қаупі ғылыми нәтижелердің тексерілуі, түсіндірілуі және жауапкершілік мәселелеріне қойылатын талаптарды күшейтеді. Сондықтан Қазақстан жағдайында ЖИ-ді ғылымға тиімді енгізу ұлттық ғылыми деректер мен қазақ тіліндегі сапалы корпустарды дамытумен, ашық әрі тексерілетін зерттеу практикасын қалыптастырумен және адамға бағдарланған құқықтық реттеумен қатар жүзеге асырылуы тиіс.

Пайдаланылған әдебиеттер:

1. Wang H., Fu T., Du Y. et al. Scientific discovery in the age of artificial intelligence // Nature. – 2023. – Vol. 620. – P. 47–60. DOI: 10.1038/s41586-023-06221-2.

2. Rapp J.T., Bremer B.J., Romero P.A. Self-driving laboratories to autonomously navigate the protein fitness landscape // Nature Chemical Engineering. – 2024. – Vol. 1. – P. 97–107. DOI: 10.1038/s44286-023-00002-4.

3. Stanford Institute for Human-Centered Artificial Intelligence. AI Index Report 2025. – Stanford University, 2025.

4. The Royal Swedish Academy of Sciences. The Nobel Prize in Chemistry 2024: Press release. – 9 October 2024.

5. Messeri L., Crockett M.J. Artificial intelligence and illusions of understanding in scientific research // Nature. – 2024. – Vol. 627. – P. 49–58. DOI: 10.1038/s41586-024-07146-0.

6. Russell S., Norvig P. Artificial Intelligence: A Modern Approach. 4th ed. – Pearson, 2021. – 1168 p.

7. Әл-Фараби. Ғылымдар классификациясы туралы трактат // Философиялық трактаттар. – Алматы: Ғылым, 1973.

8. Popper K. The Logic of Scientific Discovery. – London: Routledge, 2002. – 545 p.

9. Kuhn T.S. The Structure of Scientific Revolutions. 4th ed. – Chicago: University of Chicago Press, 2012. – 264 p.

10. Lakatos I. Falsification and the Methodology of Scientific Research Programmes // Criticism and the Growth of Knowledge / ed. I. Lakatos, A. Musgrave. – Cambridge: Cambridge University Press, 1970. – P. 91–196.

11. Jumper J., Evans R., Pritzel A. et al. Highly accurate protein structure prediction with AlphaFold // Nature. – 2021. – Vol. 596. – P. 583–589. DOI: 10.1038/s41586-021-03819-2.

12. Abramson J., Adler J., Dunger J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3 // Nature. – 2024. – Vol. 630. – P. 493–500. DOI: 10.1038/s41586-024-07487-w.

13. Terwilliger T.C., Liebschner D., Croll T.I. et al. AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination // Nature Methods. – 2024. – Vol. 21. – P. 110–116. DOI: 10.1038/s41592-023-02087-4.

14. Desai J., Watson D., Wang V., Taddeo M., Floridi L. et al. The epistemological foundations of data science: a critical review // Synthese. – 2022. – Vol. 200. – Article 469. DOI: 10.1007/s11229-022-03933-2.

15. Floridi L. A Conjecture on a Fundamental Trade-Off Between Certainty and Scope in Symbolic and Generative AI // Philosophy & Technology. – 2025. – Vol. 38. – Article 93. DOI: 10.1007/s13347-025-00927-z.

16. Searle J.R. Minds, Brains, and Programs // Behavioral and Brain Sciences. – 1980. – Vol. 3, No. 3. – P. 417–457. DOI: 10.1017/S0140525X00005756.

17. Merton R.K. The Sociology of Science: Theoretical and Empirical Investigations. – Chicago: University of Chicago Press, 1973. – 605 p.

18. UNESCO. Recommendation on the Ethics of Artificial Intelligence. – Paris: UNESCO, 2022 (adopted 23 November 2021).

19. Қазақстан Республикасы Үкіметінің 2024 жылғы 24 шілдедегі №592 қаулысы. Жасанды интеллектті дамытудың 2024–2029 жылдарға арналған тұжырымдамасы.

20. Қазақстан Республикасының 2025 жылғы 17 қарашадағы №230-VIII ҚРЗ «Жасанды интеллект туралы» Заңы (2026 жылғы өзгерістермен).

21. OECD. Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research. – Paris: OECD Publishing, 2023.

22. AlphaFold3 – why did Nature publish it without its code? // Nature. Editorial. – 22 May 2024. DOI: 10.1038/d41586-024-01463-0.

ӘОЖ 1(574):17

Мұқатаева Айзат Амантайқызы

Ф.ғ.к., педагогикалық академияның қорр.-мүшесі
Тарих және өнер кафедрасы
«Шәкәрім университеті» коммерциялық емес акционерлік қоғамы
(Семей қ, Қазақстан)

Сапарханова Мадина Төлегенқызы,
Оценова Аружан Муратхановна,
Тұрсынғазыева Ақерке Ерболатқызы
1 – курс магистранттар
Химия-биология кафедрасы
«Шәкәрім университеті» коммерциялық емес акционерлік қоғамы
(Семей қ, Қазақстан)

ШӘКӘРІМ ФИЛОСОФИЯСЫНДАҒЫ ЕРКІНДІК ЖӘНЕ МОРАЛЬДЫҚ ТАҢДАУ МӘСЕЛЕСІ

Аңдатпа: Мақалада Шәкәрім Құдайбердіұлы философиясындағы еркіндік пен моральдық таңдау мәселесі қарастырылады. Зерттеудің өзегін адамның саналы таңдау мүмкіндігі, еркіндік пен жауапкершіліктің арақатынасы, ақыл мен ұжданның моральдық шешім қабылдаудағы қызметі және тұлғаның рухани кемелденуі құрайды. Шәкәрімнің «Үш анық» еңбегі мен шығармаларындағы адам, ар, ұждан, әділет, мейірім, ынсап, ақиқат және ерік ұғымдары өзара байланыста талданады. Ойшыл үшін еркіндік сыртқы шектеулердің болмауы ғана емес, адамның өз нәпсісі мен құмарлығын ақылға бағындырып, ар алдындағы жауапкершілігін мойындай алуы. Осы тұрғыдан моральдық таңдау адамның тек қалауына емес, шешімінің салдарын түсінуіне және оны адамгершілік өлшеммен бағалауына тәуелді. Мақала теориялық сипатта орындалып, философиялық-герменевтикалық және салыстырмалы талдау тәсілдеріне сүйенеді. Нәтижесінде Шәкәрім дүниетанымында еркіндік пен мораль бір-біріне қарсы категориялар емес, өзара шарттас ұғымдар екені көрсетіледі.

Түйін сөздер: Шәкәрім Құдайбердіұлы, еркіндік, моральдық таңдау, ар ілімі, ұждан, ақыл, жауапкершілік, әділет, мейірім, адам, рухани кемелдену.

Кіріспе. Еркіндік мәселесі философия тарихындағы ең күрделі антропологиялық және этикалық мәселелердің бірі. Адамның еркіндігі туралы сұрақ оның өз әрекетін қаншалықты өзі анықтай алатынына, сыртқы жағдайлар мен ішкі қажеттіліктердің шешімге ықпалына және жасалған таңдау үшін жауап бере алуына байланысты қойылады. Еркіндік туралы мәселені моральдан бөліп қарастыру қиын, өйткені таңдау бар жерде жауапкершілік, ал жауапкершілік бар жерде құндылықтық өлшем пайда болады.

Қазақ философиялық ойында бұл мәселені терең қарастырған тұлғалардың бірі – Шәкәрім Құдайбердіұлы. Оның шығармаларында адам табиғаты, ақиқат, жан, ар, ұждан, әділет, мейірім, ынсап, еңбек және нәпсі мәселелері біртұтас дүниетанымдық жүйе ретінде көрінеді. Зерттеушілер Шәкәрімнің моральдық философиясында ар-ұжданды адамның моральдық өзін-өзі басқаруының негізгі тетігі ретінде қарастыратынын көрсетеді [1; 2].

Мақаланың мақсаты – Шәкәрім философиясындағы еркіндік пен моральдық таңдаудың мазмұнын ашып, олардың ақыл, ұждан және жауапкершілік категорияларымен байланысын көрсету. Міндеттері: еркіндік ұғымының Шәкәрімдік мағынасын айқындау; моральдық таңдау құрылымын талдау; «Үш анықтағы» рухани-этикалық идеяларды саралау; ар ілімінің қазіргі адам үшін маңызын пайымдау.

1. Шәкәрім философиясындағы адам және еркіндік мәселесіШәкәрімнің философиялық ізденістерінің өзегінде адам мәселесі тұр. Ол адамды тек табиғи немесе әлеуметтік жаратылыс ретінде емес, рухани таңдау жасай алатын және өз әрекетіне баға бере алатын субъект ретінде түсінеді. Осы ерекшелік еркіндік ұғымын оның дүниетанымындағы маңызды категориялардың біріне айналдырады. Адамның дүниедегі орны оның мүмкіндіктерімен ғана емес, сол мүмкіндіктерді қалай пайдаланатынымен анықталады.

Ойшылдың адам туралы түсінігінде табиғи құмарлықтар мен рухани-адамгершілік талап қатар жүреді. Адамда пайда, атақ, рахат, үстемдік сияқты тілектер болуы мүмкін. Бірақ тілектің болуы оны міндетті түрде орындау керек дегенді білдірмейді. Адамның ерекшелігі – өз ішіндегі ниеттердің арасынан таңдау жасай алуында. Бұл жерде ерік ұғымы шешуші маңызға ие.

Ерік еркіндігі Шәкәрім философиясында жауапкершіліктен ажырамайды. Егер адам өз ісін саналы түрде таңдауға қабілетті болмаса, моральдық жауапкершілік туралы айту қиын. Керісінше, адам өзінің шешіміне ықпал еткен жағдайларды түсіне отырып, сол жағдайда белгілі бір бағытты таңдаса, ол моральдық субъект ретінде көрінеді. Демек, еркіндік пен жауапкершілік бірін-бірі толықтырады.

Шәкәрімдік ойда сыртқы және ішкі еркіндікті ажырату маңызды. Сыртқы еркіндік адамның әрекетіне қойылған шектеулердің болмауымен байланысты. Ал ішкі еркіндік адамның өз ашуына, қорқынышына, өзімшілдігіне және уақытша пайдаға толық тәуелді болмауынан байқалады. Адам сырттай еркін болғанымен, ішкі жағынан құмарлықтың құлы болуы мүмкін. Сондықтан нағыз еркіндік адамның өзін-өзі басқаруынан көрінеді.

Осыдан «еркін адам кім?» деген сұрақ туындайды. Шәкәрімдік тұрғыдан еркін адам – тек мүмкіндігі көп адам емес, өз таңдауының моральдық салдарын түсінетін адам. Ол әрекетін сыртқы қысымға ғана сүйеніп емес, ішкі өлшем арқылы реттейді. Бұл ұстаным қазіргі мораль философиясындағы автономия ұғымымен мазмұндық жағынан үндеседі, бірақ Шәкәрім оны ар, ұят, намыс және ұждан категориялары арқылы ұлттық дүниетаным негізінде ашады.

Жеке адамның ішкі мәдениеті қоғамның жалпы адамгершілік ахуалына әсер етеді. Сондықтан қоғамды өзгерту мәселесі сыртқы институттарды өзгертумен ғана шектелмейді; адамның ниеті мен мінезін тәрбиелеу де маңызды. Осы тұрғыдан ар ілімі жеке және қоғамдық мораль арасындағы байланысты қамтамасыз ететін ішкі тетік ретінде көрінеді [2; 3].

2. Еркіндік пен қажеттілік: адамның таңдауының шекаралары

Философия тарихында еркіндік мәселесі қажеттілік ұғымымен бірге қарастырылады. Адам белгілі бір тарихи кезеңде, әлеуметтік ортада, отбасылық және мәдени жағдайда өмір сүреді. Оның білімі, экономикалық мүмкіндігі, тәрбиесі, денсаулығы және қоғамдағы орны шешімдеріне ықпал етеді. Сондықтан адамның еркіндігі абсолютті ме деген сұрақ заңды.

Шәкәрім философиясын талдау бұл сұраққа қарапайым жауап бермейді. Ойшыл сыртқы жағдайлардың маңызын жоққа шығармайды, бірақ адамның ішкі ұстанымын шешуші факторлардың бірі ретінде қарастырады. Бір жағдайдың өзінде әр адам әртүрлі әрекет етуі мүмкін, өйткені жағдай мен әрекеттің арасында ниет, ақыл, ерік және ұждан сияқты рухани тетіктер бар.

Мысалы, адам әділетсіздікке тап болғанда оны қайталау немесе оған қарсы моральдық ұстаным білдіру мүмкіндігіне ие. Бірінші жол сыртқы ортаға бейімделу болса, екіншісі ішкі құндылыққа сүйенген таңдау болуы мүмкін. Мұнда еркіндік жағдайдың жоқтығын емес, адамның сол жағдайға беретін жауабын көрсетеді.

Ойшылдың адамды толықтай сыртқы себептердің нәтижесі ретінде түсіндіруден бас тартуы маңызды. Егер адам тек ортаның қалыптастырған мінезін қайталаса, оның моральдық субъект ретіндегі ерекшелігі әлсірер еді. Адам ақыл арқылы әрекетін талдайды, ар арқылы оның дұрыстығын бағалайды. Осы қабілеттер ішкі еркіндіктің негізін құрайды.

Еркіндікке кедергі болатын факторлардың бірі – нәпсі. Нәпсі мұнда адамның өзімшілдікке, шектен тыс пайдаға, қызғанышқа, ашуға және ләззатқа тәуелділігін білдіретін моральдық ұғым ретінде қарастырылады. Адам нәпсі талабын бірден орындағанда өзін еркін сезінуі мүмкін. Бірақ бұл ұзақ мерзімде тәуелділікке айналады. Шәкәрімдік этикада еркіндік уақытша қалауды орындауда емес, оны қажет кезде тоқтата алуда көрінеді.

Демек, адам сыртқы жағдайларды әрдайым өзгерте алмайды, бірақ сол жағдайға беретін жауабын белгілі бір деңгейде таңдай алады. Еркіндік нақты өмірлік жағдайлардан бөлінбейді; ол адамның ішкі моральдық мүмкіндігімен айқындалады. Сондықтан Шәкәрім философиясындағы еркіндік абсолютті шексіздік емес, саналы өзін-өзі басқару кеңістігі.

3. Моральдық таңдау және оның құрылымы

Моральдық таңдау – бірнеше әрекет мүмкіндігінің арасынан адамгершілік өлшемге сүйеніп бірін таңдау процесі. Күнделікті өмірде таңдау қарапайым көрінгенімен, жеке пайда мен өзгенің мүддесі қақтығысқанда, ақиқат пен өтірік, әділет пен пайда, жауапкершілік пен ыңғайлылық арасында таңдау жасау қажет болғанда оның күрделілігі анық байқалады.

Шәкәрім философиясы тұрғысынан моральдық таңдауды шартты түрде бірнеше кезеңге бөлуге болады. Біріншісі – жағдайды тану. Екіншісі – ақыл арқылы ықтимал салдарды пайымдау. Үшіншісі – ішкі ниетті тексеру. Төртіншісі – ұждан өлшемімен бағалау. Бесіншісі – шешімді іске асыру және оның салдары үшін жауап беру. Бұл кезеңдер өзара байланысқан біртұтас моральдық әрекеттің бөліктері.

Ниет ұғымы ерекше маңызды. Бір әрекеттің сыртқы нәтижесі жақсы көрінгенімен, оның астарындағы ниет пайдакүнем болуы мүмкін. Керісінше, адамның мүмкіндігі шектеулі болғандықтан толық нәтиже шығара алмауы оның ниетінің адамгершілік

сипатын жоққа шығармайды. Сондықтан Шәкәрімдік этикада ниет пен әрекеттің нәтижесі қатар бағалануы керек.

Моральдық таңдауда ақылдың қызметі тек есептеу емес. Ақыл жағдайдың мәнін түсінуге, салдарды салыстыруға, жалған мен шынның арасын ажыратуға көмектеседі. Бірақ ақыл моральдық бағытсыз қалса, оны адамгершілікке қызмет ететін күш ретінде бағалау қиын. Ақылды адам өз пайдасын өзгелерден жасырудың жолын да табуы мүмкін. Сондықтан ақыл мен ұжданның үйлесімі маңызды.

Ұждан моральдық таңдаудың ішкі өлшемі ретінде адамның өзіне ғана емес, басқа адамға да қатысты жауапкершілігін күшейтеді. «Маған тиімді ме?» деген сұрақпен шектелмей, «Бұл әділетті ме?», «Біреуге зиян келтірмей ме?», «Өз ар-ұятым алдында бұл шешімді ақтай аламын ба?» деген сұрақтар шешім кеңістігін өзгертеді. Осы мағынада ұждан сыртқы бақылаусыз да әрекет ететін моральдық механизм.

Моральдық таңдаудың қайтымсыз салдары болуы мүмкін. Айтылған сөз немесе жасалған әрекет басқа адамдардың өміріне әсер етеді. Сондықтан еркіндік жауапкершілікпен толыққанда ғана моральдық мәнге ие болады. Шәкәрім философиясының маңызды түйіні – адам таңдау мүмкіндігіне ие болған сайын ар алдындағы міндеті де күшейеді.

4. «Үш анық» және еркіндіктің рухани негіздері

Шәкәрімнің философиялық жүйесін түсінуде «Үш анық» еңбегінің орны ерекше. Мұнда дүниенің мәні, адамның жаны, ар-ұждан және өмірдің мақсаты сияқты іргелі мәселелер қарастырылады. Зерттеушілер бұл еңбектің өзегінде ақиқат пен адамгершілік мәселелерінің бір-бірінен ажырамайтынын атап көрсетеді [1; 5].

«Үш анықтағы» ізденісті еркіндік мәселесімен байланыстырғанда, адамның дүниедегі орнын түсіну оның моральдық таңдауына ықпал ететінін байқауға болады. Адам өзін тек кездейсоқ материалдық тіршілік иесі ретінде қабылдаса, өмір мақсатын жеке пайдамен шектеуі мүмкін. Ал рухани жауапкершілігін мойындаған кезде әрекеттің мәні кеңейеді. Осылайша дүниетаным мен мораль бірін-бірі толықтырады.

Шәкәрім үшін ақиқатқа ұмтылу дайын жауапты қабылдау емес, іздену, салыстыру және ақылмен тексеру процесі. Бұл ұстаным интеллектуалдық еркіндікпен байланысты. Адам белгілі бір пікірді тек беделге немесе көпшіліктің сөзіне сүйеніп қабылдамай, оны ақылмен таразылауы керек. Сондықтан ой еркіндігі моральдық еркіндіктің алғышарты ретінде көрінеді.

Дегенмен интеллектуалдық еркіндік жауапкершіліксіз қарастырылмайды. Адам өз пікірінің салдарын ескеруі тиіс. Ақылдың маңызы кез келген пікірді дұрыс деп тану емес, шындықты іздеу, қателікті мойындау және моральдық тұрғыдан жетілуге ұмтылу.

«Үш анық» негізінде еркіндіктің үш қабатын пайымдауға болады: дүниені тануға ұмтылатын интеллектуалдық еркіндік; өз ниеті мен әрекетін басқаруға мүмкіндік беретін моральдық еркіндік; өмірдің мәнін іздеумен байланысты рухани еркіндік. Бұл өлшемдер бір-бірінен оқшау емес. Адам ақиқатты іздеген сайын өз әрекетіне сыни қарайды; өз әрекетіне жауапкершілікпен қараған сайын рухани дүниесін қалыптастырады.

Шәкәрімнің ұжданды ынсап, әділет және мейіріммен байланыстыруы моральдық еркіндіктің мазмұнын нақтылайды. Ынсап адамның шектен аспауын, әділет өзгенің мүддесін ескеруді, мейірім адамға жанашырлықпен қарауды білдіреді. Олардың бірлігі адамның таңдауын тек жеке мүддемен шектемей, әлеуметтік кеңістікке енгізеді [1; 3].

5. Ұждан – еркін таңдаудың ішкі моральдық өлшемі

Шәкәрім философиясындағы ең маңызды категориялардың бірі – ұждан. Зерттеушілер оны адамның моральдық өзін-өзі басқару қабілетімен, әрекетіне ішкі баға беруімен және адамгершілік міндеттерді орындауға ұмтылысымен байланыстырады [1]. Ұждан сырттан таңылатын ереже емес, адамның ішкі дүниесінде әрекет ететін моральдық өлшем.

Ұжданды еркіндік мәселесімен байланыстырудың себебі – моральдық шешім сыртқы бақылауға тәуелсіз жүзеге асуы мүмкін. Адамды ешкім бақыламай тұрған кезде де әділетсіз әрекеттен бас тарта алады. Мұндай шешімді сыртқы жаза емес, ішкі моральдық өлшем анықтайды. Демек, ұждан сыртқы бақылаудан ішкі жауапкершілікке өтуді қамтамасыз етеді.

Ұжданның құрамдары ретінде Шәкәрім ынсап, әділет және мейірімді атайды. Ынсап адамның шектен шықпауын және қажеттілігін өлшей білуін талап етеді. Әділет басқа адамдардың құқықтары мен мүдделерін ескеруге жетелейді. Мейірім моральдық әрекетке гуманистік мазмұн береді. Үшеуінің бірлігі адамды тек өз пайдасын ойлайтын субъектіден моральдық жауапкершілігі бар тұлғаға көтереді.

Ұжданды еркіндіктің шектеуі деп түсіну дұрыс емес. Шәкәрімдік түсінікте ұждан еркіндіктің сапасын арттырады. Егер адам кез келген қалауын орындауға тәуелді болса, әрекеті сырттай еркін көрінгенімен, ішкі тәуелділікке ұқсайды. Ал өз қалауын саралап, моральдық тұрғыдан дұрыс деп тапқан әрекетті саналы түрде таңдау – еркіндіктің жоғары деңгейі.

Бұл жерде еркіндік пен тәртіптің арақатынасы да өзгереді. Сыртқы тәртіп адамды мәжбүр етуі мүмкін, ал ішкі моральдық тәртіп адамның өзін-өзі реттеуінен туындайды. Шәкәрімнің ар ілімі осы ішкі тәртіпті қалыптастыруға бағытталған. Ар талаптарын орындау сыртқы еркіндікті тарылтпай, шешім қабылдау мәдениетін қалыптастырады.

Ұждан тек жеке адамның өзіне қатысты емес. Адамның таңдауы басқа адамдарға әсер етеді. Сондықтан моральдық шешімде өзгенің қадір-қасиетін мойындау маңызды. Шәкәрімнің философиялық ойындағы адамгершілік принциптер жеке адамның рухани жетілуін қоғамдағы өзара қатынастармен байланыстырады. Осы себепті ар ілімін жеке этикамен қатар әлеуметтік моральдың негізі ретінде де қарастыруға болады [3; 6].

6. Ақыл, жүрек және ерік: моральдық шешімнің ішкі тетіктері

Шәкәрім дүниетанымында адамның моральдық әрекетін бір ғана қабілет арқылы түсіндіру жеткіліксіз. Ақыл, жүрек, ерік және ұждан өзара байланыста қарастырылады. Ақыл жағдайды түсіндіреді, жүрек адамгершілік сезімді тереңдетеді, ерік шешімді іске асырады, ал ұждан әрекеттің моральдық бағытын айқындайды.

Ақыл бар, бірақ мейірім жоқ болса, білімді адамгершілікке қарсы мақсатта пайдалану қаупі бар. Мейірім бар, бірақ ақыл жеткіліксіз болса, ниет жақсы болғанымен, нақты жағдай дұрыс бағаланбауы мүмкін. Ал дұрыс шешімді түсінгенімен, ерік әлсіз болса, адам өз ұстанымын іске асыра алмайды. Сондықтан моральдық таңдау – бірнеше рухани қабілеттің бірлескен әрекеті.

Еріктің моральдық маңызы оның белгілі бір қалауды тоқтатуға немесе ұзақ мерзімді мақсат үшін қысқа мерзімді пайдадан бас тартуға мүмкіндік беруімен байланысты. Мысалы, адам өтірік айту арқылы уақытша қиындықтан құтылуы мүмкін, бірақ оның салдары сенімнің бұзылуына әкеледі. Моральдық таңдау кезінде қысқа және ұзақ мерзімді салдар салыстырылады.

Шәкәрімнің ойлау жүйесінде адамның ішкі күресі маңызды. Адам өз бойындағы жақсы және жаман бағыттардың арасынан таңдау жасайды. Мұндай таңдау бір рет қана жасалмайды, ол күнделікті өмірде қайталанады. Сондықтан моральдық кемелдену дайын күй емес, үздіксіз өзін-өзі тәрбиелеу процесі.

Өзін-өзі тәрбиелеу еркіндікпен тікелей байланысты. Адам өз мінезін өзгерте алатынын мойындағанда ғана моральдық даму мүмкін болады. Шәкәрім үшін кемшілікті «мінезім сондай» деп ақтау жеткіліксіз. Адам өз әрекетін талдап, қателігін түсініп, оны түзетуге ұмтылуы керек. Бұл – ішкі еркіндіктің практикалық көрінісі.

Осы жерде Шәкәрім философиясының қазіргі тұлғалық даму мәселелерімен үндестігі байқалады. Қазіргі адам ақпараттың көптігі, әлеуметтік қысым және жылдам шешім қабылдау жағдайында өмір сүреді. Мұндай ортада адамның өз әрекетін сыртқы әсерлерден ажыратып, ішкі құндылықтарын анықтай алуы маңызды. Ақыл мен ұжданды қатар қою моральдық рефлексияға негіз бола алады.

7. Еркіндік, жауапкершілік және моральдық автономия

Моральдық автономия адамның шешімді сыртқы мәжбүрлеуге ғана сүйенбей, өзі мойындаған моральдық принциптерге сәйкес қабылдауын білдіреді. Бұл ұғым Шәкәрімнің ар ілімімен мазмұндық тұрғыдан жақын. Адамның моральдық жетілуі оның сыртқы бақылауға тәуелді болмай, өз әрекетін ішкі ар өлшемімен реттей алуынан көрінеді.

Жауапкершілік еркіндіктің салдары ретінде түсіндіріледі. Адам таңдау мүмкіндігіне ие болған сайын шешіміне жауап беру міндеті де артады. Шәкәрім философиясында жауапкершілік тек заң алдындағы жауапкершілік емес. Заң рұқсат еткен әрекет моральдық тұрғыдан әрдайым дұрыс болмауы мүмкін. Сондықтан ар алдындағы жауапкершілік заңдық норманы толықтыратын ішкі өлшем қызметін атқарады.

Бір адамның құпиясын білу оны таратуға мүмкіндік береді, бірақ мүмкіндік пен моральдық рұқсат бір нәрсе емес. Біреудің әлсіздігін пайдалану тиімді болуы мүмкін, бірақ әділет пен мейірім өлшеміне сәйкес келмеуі ықтимал. Мұндай жағдайда адамның шынайы еркіндігі мүмкіндіктерінің көптігінде емес, сол мүмкіндіктердің қайсысын моральдық тұрғыдан таңдағанында көрінеді.

Шәкәрім философиясында адам өзіне ғана есеп бермейді, оның әрекеті қоғамға әсер етеді. Сондықтан жеке мораль қоғамдық өмірдің сапасына ықпал етеді. Әр адам өз таңдауында өтірікке, әділетсіздікке және қиянатқа жол бермеуге тырысса, бұл жеке деңгейдегі ұстанымнан қоғамдық құндылыққа айналады.

Ойшылдың ар ілімі моральдық автономияның ұлттық-философиялық үлгісі ретінде қарастырылуы мүмкін. Ол дайын моральдық ережелерді жаттаудан гөрі адамның ішкі сенімін қалыптастыруға басымдық береді. Адам «неге бұлай істеуім керек?» деген сұраққа ақылы мен ұжданы арқылы жауап іздегенде моральдық шешім шынайы ішкі таңдау сипатына ие болады.

Моральдық автономияны абсолютті субъективизммен теңестіруге болмайды. «Мен осылай ойлаймын» деген жеке пікір кез келген әрекетті ақтай алмайды. Шәкәрім үшін таңдау ақиқат, әділет, мейірім және ұждан сияқты жалпыадамзаттық өлшемдермен байланыста қарастырылады. Демек, еркіндік субъективті қалау мен моральдық талап арасында саналы тепе-теңдік орнатуды қажет етеді.

8. Шәкәрімнің еркіндік философиясының қазіргі маңызы

Шәкәрімнің еркіндік пен моральдық таңдау туралы ойлары қазіргі қоғам үшін де өзекті. Қазіргі кезеңде адамның таңдауы көбейгенімен, шешіміне ықпал ететін факторлар да артты. Ақпараттық орта, әлеуметтік желілер, қоғамдық пікір, жарнама және тұтыну мәдениеті адамның қалауы мен мінез-құлқына әсер етеді. Сыртқы еркіндіктің кеңеюі автоматты түрде ішкі еркіндіктің күшеюін білдірмейді.

Осындай жағдайда ішкі моральдық бақылауға негізделген ар ілімі маңызды. Адам ақпаратты бірден қабылдамай, оның шынайылығын ақылмен тексеруі; эмоциямен қабылданған шешімді қайта бағалауы; өз әрекетінің басқа адамдарға әсерін ескеруі қажет. Бұл принциптер адамның ақпараттық және әлеуметтік ортадағы моральдық дербестігін күшейтеді.

Білім беру саласында да бұл идеяның маңызы бар. Моральдық тәрбиені тек ережелерді жаттатумен шектеу оқушының ішкі жауапкершілігін толық қалыптастырмайды. Моральдық білім нақты жағдайларда таңдау жасауға мүмкіндік беретін рефлексиялық қабілетті дамытуы тиіс. «Егер сен осындай жағдайда болсаң, не істер едің?», «Бұл шешім кімге әсер етеді?», «Өз шешіміңді ар алдында қалай түсіндіресің?» сияқты сұрақтар моральдық ойлауды дамыта алады.

Шәкәрімнің еркіндік туралы көзқарасын азаматтық мәдениет тұрғысынан да талдауға болады. Еркін қоғамда адамның өз пікірін айту құқығы бар, бірақ бұл құқық өзге адамның қадір-қасиетін жоққа шығаруға негіз болмауы керек. Өзгенің еркіндігін мойындау – өз еркіндігіңнің моральдық шекарасын түсіну. Осы тұрғыдан еркіндік пен өзара құрмет бір-біріне қарсы емес.

Шәкәрім философиясының тағы бір қазіргі маңызы – жауапкершілікті сыртқы институттарға ғана жүктемеу. Қоғамдық тәртіпті заңдар қамтамасыз етеді, бірақ олардың бәрін бақылау мүмкін емес. Адамның ішкі ұстанымы әлсіз болса, тәртіп тек бақылау кезінде сақталуы мүмкін. Ал ұждан қалыптасқан жағдайда адам бақылаусыз жерде де моральдық шешім қабылдай алады.

Сондықтан Шәкәрімнің ар ілімін қазіргі этикалық білім, тұлғалық даму және азаматтық жауапкершілік мәселелерімен байланыстыра зерттеу мүмкіндігі кең. Оның философиясы еркіндікті адамның тек құқықтары арқылы емес, өз әрекетінің салдарын түсіну қабілеті арқылы пайымдауға мүмкіндік береді.

9. Еркіндік пен моральдық таңдау категорияларының өзара байланысы

Шәкәрім философиясын жинақтап қарастырғанда еркіндік пен моральдық таңдаудың бірнеше деңгейде байланысатынын көреміз. Бірінші деңгей – танымдық деңгей. Адам не істеп жатқанын түсінбесе, оның таңдауы толық саналы болмайды. Сондықтан еркіндік ақиқатқа ұмтылумен және ақылдың қызметімен байланысты. Екінші деңгей – еріктік деңгей. Адам дұрыс деп таныған әрекетін іске асыру үшін ерік күшін жұмсайды. Үшінші деңгей – моральдық деңгей. Таңдаудың дұрыстығы ұждан, әділет және мейірім өлшемдерімен бағаланады.

Бұл құрылымда моральдық таңдау тек шешім қабылдау актісі емес, адамның өзін қалыптастыратын процесс. Әр таңдау адамның мінезіне әсер етеді. Әділетті әрекетті қайталау әділеттілікке бейімділікті күшейтеді, ал өтірікті әдетке айналдыру адамның моральдық сезімін әлсіретуі мүмкін. Осы тұрғыдан еркіндік адамның болашақ тұлғалық сипатын да қалыптастырады.

Шәкәрімнің ойынша адам өзінің ішкі дүниесін өзгертуге қабілетті. Бұл мүмкіндік моральдық тәрбиенің теориялық негізін құрайды. Егер адам өзгермейтін, толық

детерминацияланған тіршілік иесі болса, тәрбие мен өзін-өзі жетілдірудің мағынасы азаяр еді. Ал адамның таңдауға қабілеттілігі оның рухани дамуына жол ашады.

Еркіндіктің моральдық мазмұны адамның өзгені мойындауынан да көрінеді. Өз еркіндігін абсолюттендірген адам басқа адамның еркіндігіне қауіп төндіруі мүмкін. Шәкәрімнің әділет және мейірім категориялары осы қарама-қайшылықты жұмсартады. Әділет теңдік пен құқықтық өлшемді, мейірім адамаралық қатынастың гуманистік өлшемін қамтамасыз етеді.

Моральдық таңдау кезінде адамның ішкі мотиві мен қоғамдық салдары қатар қарастырылуы қажет. Шәкәрімнің философиясы осы екі деңгейді бір-бірінен ажыратпайды. Адамның жүрегіндегі ниет сыртқы әрекет арқылы қоғамға әсер етеді, ал қоғамдағы моральдық ахуал жеке адамның қалыптасуына қайта ықпал етеді. Бұл өзара байланыс жеке және қоғамдық этиканың тұтастығын көрсетеді.

Осыдан Шәкәрім философиясындағы еркіндікті үш сөзбен түйіндеуге болады: таңдау, ұждан, жауапкершілік. Таңдау еркіндіктің мүмкіндігін білдіреді; ұждан оның моральдық бағытын анықтайды; жауапкершілік сол таңдаудың салдарын мойындауды қамтамасыз етеді. Үшеуі біріктірілгенде ғана еркіндік адамгершілік сипат алады.

Қорытынды

Шәкәрім Құдайбердіұлының философиялық мұрасындағы еркіндік пен моральдық таңдау мәселесін талдау адамды рухани және моральдық субъект ретінде түсінудің маңызын көрсетеді. Ойшыл үшін адам сыртқы жағдайлардың пассивті нәтижесі емес, өз әрекетін ой елегінен өткізіп, оған моральдық баға бере алатын тұлға. Сондықтан еркіндік оның философиясында жауапкершілікпен тікелей байланыста қарастырылады.

Шәкәрімдік түсініктегі еркіндік «қалағаныңның бәрін жасау» қағидасына сыймайды. Еркіндік адамның өз қалауын тануы, оны ақылмен саралауы, ұжданмен өлшеуі және саналы шешім қабылдауы арқылы жүзеге асады. Бұл жерде ішкі еркіндік сыртқы еркіндікпен салыстырғанда ерекше маңызға ие. Адам сыртқы қысымнан бос болғанымен, нәпсі мен өзімшілдікке тәуелді болса, оның еркіндігі толыққанды емес.

Моральдық таңдау осы ішкі еркіндіктің нақты көрінісі. Адам алдында бірнеше мүмкіндік тұрғанда оның шешімі ниет, ақыл, ерік және ұжданның өзара әрекеті арқылы қалыптасады. Ақыл әрекеттің салдарын түсіндіреді, мейірім басқа адамның жағдайын ескеруге мүмкіндік береді, ерік шешімді іске асырады, ал ұждан оның моральдық дұрыстығын бағалайды.

«Үш анық» пен ар ілімін талдау Шәкәрім философиясында танымдық және моральдық мәселелердің біртұтас екенін көрсетеді. Ақиқатты іздеу адамның өзін және әрекетін тануына алып келеді; өзін тану жауапкершілікті күшейтеді; жауапкершілік рухани кемелденуге жағдай жасайды. Сондықтан еркіндік, ақыл, ұждан және рухани даму өзара байланысқан ұғымдар жүйесін құрайды.

Шәкәрімнің ұжданды ынсап, әділет және мейіріммен байланыстыруы еркіндіктің моральдық мазмұнын айқындайды. Адам өз еркіндігін басқа адамның қадір-қасиетін бұзбай жүзеге асыруы тиіс. Осы тұрғыдан еркіндік пен жауапкершілік бір-бірін шектейтін емес, бір-бірін мәнді ететін категориялар. Еркіндік таңдау мүмкіндігін берсе, жауапкершілік сол таңдауға моральдық өлшем береді.

Зерттеу нәтижесінде Шәкәрім философиясындағы моральдық таңдау мәселесін қазіргі қоғамның ақпараттық, әлеуметтік және мәдени жағдайларымен байланыстыра қарастыруға болатыны көрсетілді. Қазіргі адамда таңдау мүмкіндігі көп болғанымен,

сыртқы ықпалдар да көп. Сондықтан ішкі еркіндік пен моральдық автономияны дамыту өзекті. Шәкәрімнің ар ілімі осы мәселені адамның ішкі дүниесін тәрбиелеу, ақыл мен ұжданды үйлестіру және өзгенің еркіндігін мойындау арқылы пайымдауға мүмкіндік береді.

Қорытындылай келе, Шәкәрім философиясында еркіндік пен моральдық таңдау бір-бірінен бөлек қарастырылмайды. Адамның шынайы еркіндігі оның таңдауының адамгершілік мазмұнымен анықталады. Ал моральдық таңдау адамның өз әрекеті үшін жауапкершілік ала алуымен толықтырылады. Осы тұрғыдан Шәкәрімнің философиялық мұрасы қазақ ойлау кеңістігіндегі моральдық автономия, ар-ұждан және жауапкершілік мәселелерін зерттеу үшін маңызды теориялық негіз болып табылады.

Пайдаланылған әдебиеттер

1. Асқарова А.Ш., Баязитов Б.Б. Шәкәрімнің «Ар ілімі» концепциясы // Eurasian Journal of Philology: Science and Education. – 2019. – №2 (174). – Б. 19–26. DOI: 10.26577/EJPh.2019.v174.i2.ph3.

2. Шәкәрім дүниетанымындағы адам мәселесі // Journal of Philosophy, Culture and Political Science. – 2014. – №2 (47). – Б. 3–11.

3. Шәкәрім Құдайбердіұлының моральдық философиясы мен ұждан тұжырымдамасы. – 2021.

4. Шәкәрім Құдайбердіұлы. Үш анық. – Алматы: Қазақстан, 1991.

5. Шәкәрім Құдайбердіұлы. Шығармалары: өлеңдер, дастандар, қара сөздер, философиялық трактаттар. – Алматы: Жазушы, 1988.

6. Шәкәрім Құдайбердіұлы. Өлеңдер мен поэмалар. – Алматы: Раритет, 2006.

7. Қазақ философиясының тарихы: он бес томдық. – Алматы: ҚР БҒМ ҒК Философия, саясаттану және дінтану институты, 2014.

8. Нысанбаев Ә. Қазақ философиясы: тарихы мен қазіргі мәселелері. – Алматы: Ғылым, 1999.

9. Әл-Фараби. Әлеуметтік-этикалық трактаттар. – Алматы: Ғылым, 1975.

10. Шәкәрімнің ар-ұждан философиясы және тұлғаның моральдық қалыптасуы // Шәкәрім университетінің ғылыми басылымдары. – 2021. – №10 (134).

UDC 81'27

Komekova Saida
Grade 11 student
Scientific supervisor: Dyussembina Kalamkas Serikovna
School-Lyceum No. 120 named after K. Satbayev
(Astana, Kazakhstan)

DO UNHAPPY REVIEWERS MIX IN MORE RUSSIAN? STAR RATINGS AND KAZAKH-RUSSIAN CODE-SWITCHING IN 175,002 ONLINE REVIEWS

Abstract: Bilingual people often say that they switch languages when they are emotional, but no measurement of this for Kazakh and Russian was found. This study asks whether Kazakh-speaking reviewers mix Russian words into their reviews more often when they are dissatisfied (1-2 stars) than when they are satisfied (4-5 stars). The data are all 175,002 non-empty reviews of the public KazSAnDRA dataset of Kazakh online reviews, each with a star rating from 1 to 5. For every review, the share of Russian words among all words was counted with an automatic rule that the author's mentoring team built from a boundary and twenty example words she wrote before the count: an old, established loanword such as *телефон* or *автобус* counts as Kazakh, while a Russian word used instead of an existing Kazakh one, such as *скидка* or *доставка*, counts as Russian. A word-by-word check of all 1,763 words in 160 randomly drawn reviews (60 of them drawn from reviews the rule marked as mixed), made by an AI assistant, found 90% precision and 83% recall (92% and 83% in the fully random part). The mean share of Russian words falls with the rating, from 3.85% in 1-star reviews to 2.49% in 5-star reviews (ratio 1.55, 95% CI 1.50 to 1.60), and 1-2-star reviews contain 1.49 times as much Russian as 4-5-star reviews. The direction predicted by the author's hypothesis is confirmed, but her forecast of "three times" was about twice the observed ratio. The difference comes almost entirely from how often a review contains any Russian (23.5% of 1-star against 15.5% of 5-star reviews); among reviews that do, the share of Russian words is only slightly higher (1.05 times in short and 1.18 times in long reviews). The difference holds in every additional check (smallest ratio 1.17, at 10 per million) and is larger in long reviews (1.79 times) than in short ones (1.43). Relative to their frequency, the largest excess (3.15 times; a grouping made after a first review) is in Russian emphatic particles and degree adverbs such as *вообще*, *уже*, *еще* and *даже*; they are about a tenth of the extra Russian words, most of which are content words such as *реклама* and *обновление*.

Keywords: code-switching; Kazakh-Russian bilingualism; emotion; sentiment; online reviews; star ratings; corpus linguistics; KazSAnDRA.

1. Introduction

In Kazakhstan many people use two languages, Kazakh and Russian, and move between them inside one conversation or even one sentence. The author's first study looked at how this mixing is built: in 2,447 Kazakh sentences with Russian words from online reviews, almost every switch was a single Russian word inserted into a sentence whose grammar stayed Kazakh [1]. That study answered the question of how people mix. It did not ask when they mix.

One common answer to "when" is "when we are emotional": the author's own observation, written before the count (Section 3), was that people switch to Russian to express criticism, and code-switching is described as a strategy for expressing emotion [8]. Kazakh and

Russian are the two main languages of Kazakhstan, and how each of them is used has been studied for years [2], but no test of the link between emotion and mixing on real written Kazakh was found. The online reviews used in the first study offer a chance to make one, because every review in the KazSAnDRA dataset carries a rating from 1 to 5 stars, which the dataset's creators describe as a measurable representation of the reviewers' attitudes [3]. The rating is therefore a direct signal of how satisfied the writer was.

This paper therefore asks one narrow question: do people mix Russian into Kazakh more often in dissatisfied reviews (1-2 stars) than in satisfied reviews (4-5 stars)? It uses all 175,002 non-empty reviews of the public dataset and measures mixing as the share of Russian words in each review. The author's hypothesis and a numerical forecast were written down before any count was made.

2. Background

2.1 Language choice and emotion in bilinguals

Research on bilingualism shows that the two languages of a bilingual person do not carry the same emotional weight. In a survey of 1,039 multilinguals, swearwords and taboo words felt most powerful in the first language [4]. In about 430,000 tweets of Hindi-English bilinguals, Hindi, the native language, was preferred for negative opinions and swearing [5]. But in a Spanish/Galician conversation, indignation was expressed in Spanish, the dominant language, and contempt in Galician, the less dominant one [6]. Switching languages was also tied to emotion and emotional labour for 15 Iranian English teachers in class [7], and a review in this journal describes code-switching as one of the strategies bilinguals use to express emotion [8]. Because the direction differs between pairs, it has to be measured for each pair.

2.2 Measuring emotion in mixed-language text

Large-scale work on emotion in mixed text comes mostly from computational linguistics, where annotators decide the emotion of each text. Lee and Wang annotated Chinese-English posts from the social network Weibo for the emotion each language expresses and found English used more for happiness than for sadness [9], and Tang and colleagues fine-tuned BERT classifiers to detect several emotions at once in Chinese-English text [10]. Annotators often disagree: in the English-Spanish EN-ES-CS corpus of Vilares and colleagues, Krippendorff's alpha was 0.629 to 0.664 for negative and 0.500 to 0.693 for positive sentiment [11]. Star ratings avoid this problem because the writer gives the rating, but ratings on review platforms usually form a "J-shaped" distribution, with many 5-star ratings, some 1-star ratings and few in between [12]; KazSAnDRA has this shape, with 131,021 of its 180,064 reviews at 5 stars [3].

2.3 The gap

Three papers in this journal discuss emotion or language mixing without counting it in texts: a review of emotional expression in two languages [8], a description of youth slang and language mixing on social networks [13] and a survey-based study of language, psyche and society [14]. Kazakh-Russian code-switching was counted in real reviews in the author's first study [1] and tested on language models in a benchmark [15], but neither uses an emotion signal. The present study joins the two lines.

3. Hypothesis and forecast

The research question was fixed on 17 September 2026. The next day, before any count, the author wrote the hypothesis (her texts are edited for grammar and spelling; words added or replaced for meaning are in square brackets):

"I hypothesize that there would be more Russian words in negative reviews [rated] 1-2 than in positive reviews [rated] 4-5. Perhaps people switch to Russian language in negative

feedback in order to emphasize their dissatisfaction or highlight the issue emotionally. In positive comments, on the other hand, language can stay more plain since the author does not need to emphasize emotions that intensively. Therefore, I anticipate more frequent Kazakh-Russian language mixing in negative comments specifically."

On 4 October 2026, again before any count, she added the reason for her forecast:

"I suppose that there would be more Russian words in reviews with 1-2 stars specifically. According to my observations, people tend to switch to Russian in cases when they want to express their criticism or dissatisfaction. Kazakh language, on the other hand, is used in a laconic style when the person is satisfied. The reason for this difference lies in the fact that Russian has a more extensive critical vocabulary, which is needed for expressing discontent. In Kazakh, thank-you lexis prevails."

Asked on 5 October how many times higher the share of Russian words in 1-2-star reviews would be than in 5-star reviews, with "one and a half or three times" given as examples in the question, she answered: three times. This is the forecast checked in Section 6.2. The analysis plan (Section 5.4) was written by her mentor and approved by the author on 6 October 2026, also before any share was computed.

4. Data

4.1 The KazSAnDRA dataset

On 2 October 2026 the author summarised the dataset paper [3]: the reviews come from four sources, namely an app store for Android devices (Appstore), an online library of Kazakh books and audiobooks (Bookstore), map and navigation services (Mapping) and online marketplaces (Market); 180,064 reviews were collected, of which 25,235 have one star, 4,929 two, 7,262 three, 11,617 four and 131,021 five; and the dataset is released under the CC BY 4.0 licence.

According to the dataset paper, the reviews were collected between September 2022 and September 2023, checked by native speakers of Kazakh and left as written [3], so they keep their misspellings, Kazakh typed in Russian letters, Latin script and Russian words; the release keeps this original text (column text) next to the lower-cased copy without punctuation and emoji (text_cleaned) that the dataset paper describes as its pre-processing [3, 16], and only the original text is used here.

4.2 Which reviews were used

The dataset paper reports 180,064 collected reviews [3]. Before publishing, its creators removed duplicates (identical text and score) [16]; the public release contains 175,158 reviews with their original 1-5 scores (the three "score classification" files: 140,126 training, 17,516 validation and 17,516 test reviews) and 173,374 distinct review texts. The author's first study reported 172,015 reviews [1]. A recount made for this paper shows that 172,015 is exactly the number of distinct review texts in the polarity files plus the score-classification training file, that is, without the score-classification validation and test files, so the first study most likely scanned about 99% of the distinct review texts. This paper uses the complete release: 175,158 reviews, of which 156 contain no words at all (only emoji, digits or punctuation), leaving 175,002 reviews with 1,507,052 words.

4.3 Why star ratings, and their weak points

The author explained the choice of star ratings as follows (24 September 2026):

"The usage of authors' ratings (on a scale from 1 to 5 stars) as a metric of the emotional sentiment of the review has a principal advantage in comparison with traditional expert annotation, since the author's rating measures emotions more faithfully and more accurately

than ratings assigned by external annotators. In conventional corpora (for example, the EN-ES-CS Corpus) the sentiment annotations of external experts demonstrate a low level of agreement (inter-annotator agreement), as people rate someone else's anger or approval, filtering the text through their own perception. Ratings provided by the authors themselves reflect their authentic internal state during writing. The author is not trying to 'guess' the emotion: they show it, [which] eliminates the misrepresentations inevitable in external interpretations of textual markers. Nonetheless, this approach has its own weaknesses that have to be considered in the analysis. Scale subjectivity: for one user, deep disappointment (high level sentiment) is expressed as 1 star, while the other one puts 3 stars with [a similar] anger level ('it could be worse'). Cognitive shift: users tend to write reviews in an 'affect', that is, emotional excitement. For this reason, the selection naturally shifts to polar values (1 or 5 stars), diluting the neutral zone. Textual discrepancy: a star rating measures the overall impression of an experience, whereas the review itself may contain complex code-switching, sarcasm or a detailed description of specific advantages and disadvantages, which formally contradict the overall rating."

The agreement measured there was low to moderate [11]. This study treats the stars as a measure of satisfaction rather than of emotion as such (Section 8).

These weak points are visible in the data. Only 7,184 of the 175,002 reviews (4.1%) have 3 stars, while 126,525 (72.3%) have 5 stars. For this reason the study compares shares inside each rating group, never raw numbers of Russian words, and it reports the planned comparison of 1-2 against 4-5 stars together with the comparison of the two extremes, 1 against 5 stars.

5. Method

All counts and statistics in this paper were computed in Python by the author's mentoring team with an AI assistant (see the contributions note at the end).

5.1 What counts as a word

A word is a run of letters (Cyrillic or Latin; parts joined by a hyphen are kept together) in the original review text, written in lower case. Digits, emoji and punctuation are not words. By this definition the median review has 6 words and the mean review 8.6 words.

5.2 What counts as a Russian word

Many Russian-origin words have been part of Kazakh for decades, so "Russian origin" cannot be the test. The boundary used here is the author's own, written on 27 September 2026, before the count:

"I would draw a line depending not on the word origin, but on the perception of this word as common in Kazakh speech. If the word has been used in Kazakh for a long time, has a usual form, and appears in daily speech without feeling like a switch, I would not count it as code-switching. If the word looks like Russian but is used instead of a Kazakh one, I count it as a Russian insertion. For example, I do not consider these words as a [switch]: «Телефон», «Машина», «Магазин», «Компьютер», «Автобус», «Завод», «Банк», «Билет», «Паспорт», «Телевизор». These words, on the other hand, are Russian insertions for me: «Скидка», «Доставка», «Заказала», «Понравилось», «Качество», «Обслуживание», «Удобно», «Дорого», «Продавец», «Размер». This line is needed to count cases of switching to Russian specifically, but not the words that are normally used in Kazakh."

There are 141,940 different words in the reviews, too many to judge one by one, so the mentoring team turned the rule into steps that a computer applies to every word:

1. A word in Latin letters is not Russian.

2. The author's ten Kazakh words above, with or without Kazakh endings (*телефонды*, *банкке*), are not Russian; her ten Russian words are Russian.

3. Russian function words that have no common Kazakh twin (*и*, *в*, *с*, *что*, *это*, *очень*, *уже* and similar; the list was drawn up by the AI assistant) are Russian; two of them, *я* and *о*, can also be Kazakh particles and make up 1.3% of the Russian words counted. Words that exist in both languages, such as *да* and *не* (Kazakh "too" and "what"), *он* (Kazakh "ten") or *а*, are not counted as Russian.

4. A Russian stem with Kazakh endings (*приложение-ге*, *доставка-сы*, *курьер-ге*) is counted as a Russian word, because the word itself is Russian and is used instead of a Kazakh one (*қолданба*, *жеткізу*); only its ending is Kazakh. Such words make up 17.5% of all Russian words found. This is more than the 4.9% of single-word switches with Kazakh endings in the first study [1]; the two figures come from different counting rules and units and are not directly comparable.

5. Any other word is Russian if it is in the Russian dictionary of the morphological analyser pymorphy3 (based on pymorphy2 [17]; proper names excluded) and it is not established in Kazakh. "Established" is measured in edited Kazakh text: the word and its Kazakh-suffixed forms must occur at least 1.75 times per million words in Kazakh Wikipedia [18] (22.2 million words, counting only clearly Kazakh sentences), also when Kazakh letters are typed as Russian ones (*көп* typed as *коп*).

The threshold of 1.75 per million was set from the author's twenty words. In Kazakh Wikipedia, her ten "Russian" words occur at most 1.62 times per million (*размер*), and her ten "Kazakh" words at least 1.85 times per million (*магазин*, one of the four examples her mentor gave when he asked for the list); any threshold between the two reproduces all twenty of her decisions. Finally, the AI assistant reviewed the 300 most frequent words that the steps marked Russian and the 150 most frequent mixed words, reading uncertain ones in context, and made 46 corrections: 18 Kazakh words typed in an unusual way (for example *дым* "nothing at all", *кермет* for *керемет*), 27 names and interjections (for example the brand *Флип* and *вау*) and one Kazakh particle (*ко*). These corrections were made on word frequencies over all reviews, without looking at the star ratings.

Table 1 shows the rule on two reviews drawn at random from the data (one 5-star and one 1-2-star review of 8 to 30 words containing at least one Russian word, without obscene words).

Table 1. The counting rule on two randomly drawn reviews

Review (stars)	Words	Russian words	Share
«Өте керемет еш минус жоқ мен сол себептен 5 звёзд қойып отырм» (5)	11	1 (звёзд)	9.1%
«Қазақ тілі жоқ. Қазақ тілді қауымның құқығын шектеп тұрсыз! Нету казахского языка(» (1)	12	3 (нету, казахского, языка)	25.0%

The rule does not count *минус* (1.85 per million in Kazakh Wikipedia, just above the threshold); the check in Section 5.5 judged a use of *минус* in place of Kazakh *кемшілік* to be

Russian, so words like this are among the rule's misses. *Звёзд* is a Russian case form. The second review is a complaint that an application has no Kazakh language, written first in Kazakh and then repeated in Russian.

5.3 The measure: the share of Russian words

For each review the measure is the number of Russian words divided by the number of all words. Her mentor set out two options, a simple "yes or no" and the share, with the example of a 100-word and a 10-word review; the author chose the share over the other option her mentor offered and explained why (25 September 2026):

"For every review I count the proportion of Russian words among all words. I think this method is more accurate because it shows not only the presence of [mixing], but also how conspicuous it is in the commentary itself. For example, two Russian words in a 100-word review and one Russian word in a 10-word review are not the same thing, although with 'yes or no' both reviews would simply be considered mixed. Therefore, the 'yes/no' option is too harsh: it does not take into account the length of the feedback and the number of Russian words. The percentage allows [us] to compare reviews of different sizes and see the real degree of mixing."

The "yes or no" measure is still reported as a check (Section 6.4).

5.4 The planned comparisons

The plan approved on 6 October 2026 was: (1) compute the share of Russian words in every review and the mean share for each rating; (2) compare 1 against 5 stars and 1-2 against 4-5 stars, comparing shares, not numbers of words, because the groups differ in size; (3) repeat the comparison separately for short reviews (10 words or fewer) and long reviews (more than 10 words), so that a difference cannot simply be a difference in length. The forecast is checked by the ratio of the mean share in 1-2-star reviews to the mean share in 5-star reviews.

For each comparison the paper reports the two mean shares, their difference and their ratio with 95% bootstrap confidence intervals (2,000 resamples of reviews within each group), the p-value of the Mann-Whitney U test, and the rank-biserial correlation as an effect size (0 means no difference, 1 means every review of one group is above every review of the other). The confidence intervals of single means (Table 2, Figure 1) use the normal approximation. With groups of tens of thousands of reviews even tiny differences have tiny p-values, so the paper relies on the size of the difference.

Checks added after the plan are reported as such: domains; the "yes or no" measure; a context version of the rule, in which a shared word (such as *не*) counts as Russian when the nearest non-shared word on each side, where one exists, is Russian; only reviews with a Kazakh-specific letter; thresholds of 1, 2.5, 5 and 10 per million; texts that occur once; and, after a first review of this paper, the share among mixed reviews, the classes of Table 5 and runs of Russian words.

5.5 Word-by-word check of the rule

The rule was checked on two random samples, with the star ratings hidden: 100 reviews drawn from all reviews with at least 3 words (913 words), and 60 reviews drawn from the reviews in which the rule found at least one Russian word (850 words). Each word was read against the author's boundary in Section 5.2. The reading was done by the same AI assistant that wrote the code, so it is a second application of the boundary, not an independent human judgement. Loanwords under 2.5 per million in Kazakh Wikipedia (the next threshold tested), here *минус*, *пароль*, *реанимация* and *ресторан*, were judged Russian when used in place of a Kazakh word (*кемшілік*, *құпиясөз*, *жансақтау*, *мейрамхана*), and a shared word such as *а* or *не* was judged Russian inside a Russian phrase. Over the 1,763 words, the rule agreed with the

reading on 98.0% of words. Of the words the rule marked Russian, 89.9% were Russian (precision), and it found 83.5% of the Russian words (recall). In the fully random sample the figures were 92.1% and 83.3%. The missed words were mostly misspelt Russian words (*саглашусь*, *спаибо*) and slang, and the false alarms were mostly names and Kazakh endings typed apart from their word (*мб ты*). When the ratings were joined after the reading, the errors were not tied to the rating: in 1-2-star reviews precision was 0.94 and recall 0.87, in 4-5-star reviews 0.85 and 0.79, and in both groups the rule counted 0.93 times as many Russian words as the reading.

6. Results

6.1 Russian words by rating

Across all 175,002 reviews, 47,028 of 1,507,052 words (3.1%) are Russian, and 17.3% of reviews contain at least one Russian word. Reviews written wholly in Russian are almost absent (50 reviews), as expected in a dataset collected as reviews in Kazakh [3]; the Russian words appear inside Kazakh reviews, as in the first study [1].

Table 2. Russian words by star rating

Stars	Reviews	Words	Russian words	Reviews with any Russian, %	Mean share, % (95% CI)
1	25,019	232,629	10,795	23.5	3.85 (3.74 to 3.97)
2	4,896	44,644	1,756	21.1	3.50 (3.24 to 3.75)
3	7,184	67,842	2,654	21.8	3.58 (3.38 to 3.78)
4	11,378	102,838	3,453	19.5	3.09 (2.94 to 3.24)
5	126,525	1,059,099	28,370	15.5	2.49 (2.45 to 2.53)

The share of Russian words is highest in 1-star reviews and lowest in 5-star reviews, and it falls almost step by step in between (Table 2, Figure 1a); only 2 and 3 stars are in reverse order, and their confidence intervals overlap.

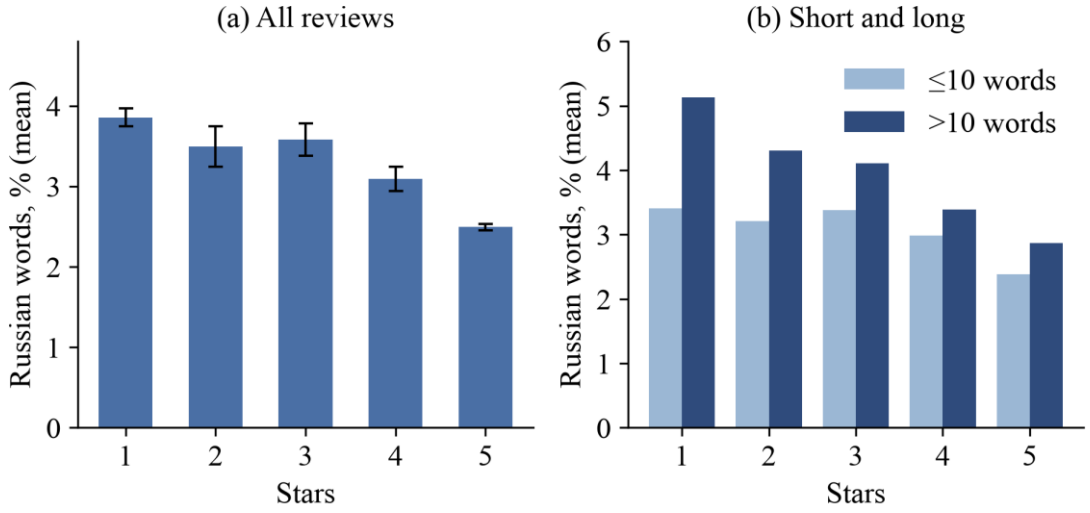


Figure 1. Mean share of Russian words by star rating: (a) all reviews, with 95% confidence intervals; (b) short reviews (10 words or fewer) and long reviews (more than 10 words)

6.2 The planned comparisons and the forecast

Table 3. Planned comparisons of the mean share of Russian words

Comparison	Reviews	Mean share, %	Difference, percentage points (95% CI)	Ratio (95% CI)	Rank-biserial r
1 vs 5 stars	25,019 vs 126,525	3.85 vs 2.49	1.36 (1.25 to 1.48)	1.55 (1.50 to 1.60)	0.08
1-2 vs 4-5 stars	29,915 vs 137,903	3.80 vs 2.54	1.25 (1.14 to 1.36)	1.49 (1.45 to 1.54)	0.07
1-2 vs 5 stars (forecast)	29,915 vs 126,525	3.80 vs 2.49	1.30 (1.19 to 1.42)	1.52 (1.47 to 1.57)	0.08

All three differences are in the direction of the hypothesis, with Mann-Whitney $p < 0.001$. Dissatisfied reviews contain about one and a half times as much Russian as satisfied reviews. The forecast was three times: the observed ratio for 1-2 against 5 stars is 1.52, and the whole confidence interval (1.47 to 1.57) lies far below 3, so the forecast was about twice the observed ratio. The hypothesis is supported in direction, not in size.

The effect is small in the sense of the rank-biserial correlation (0.07 to 0.08), because most reviews in every group contain no Russian at all. The difference comes almost entirely from how often a review contains any Russian (23.5% of 1-star reviews against 15.5% of 5-star reviews). Among the 30,249 reviews that do contain Russian, the mean share of Russian words is almost the same: 16.41% for 1 star against 16.12% for 5 stars (ratio 1.02, 95% CI 1.00 to 1.04), and 16.44% for 1-2 stars against 16.10% for 4-5 stars (ratio 1.02, 95% CI 1.00 to 1.04). It is higher in both length groups (1 against 5 stars, long: 1.18, 95% CI 1.15 to 1.22; short: 1.05, 1.02 to 1.07); the overall ratio is near 1 because mixed 1-star reviews are more often long (53.7% against 45.5%), and long reviews have lower shares (9.5% against 22.4% in short mixed reviews).

6.3 The length check

Short reviews (10 words or fewer) make up 76.7% of the data (134,289 reviews). The difference holds in both length groups (Figure 1b): in short reviews the share is 3.40% for 1 star against 2.39% for 5 stars (ratio 1.43, 95% CI 1.36 to 1.50), and in long reviews it is 5.13% against 2.87% (ratio 1.79, 95% CI 1.72 to 1.87). For 1-2 against 4-5 stars, the ratios are 1.39 (1.33 to 1.45) for short and 1.71 (1.64 to 1.78) for long reviews. The difference is not produced by length; it is larger in long reviews, perhaps because the writer has room to explain the problem (not tested).

6.4 Additional checks

Table 4. Additional checks (not in the plan): ratio of mean shares (95% CI)

Check	1 vs 5 stars	1-2 vs 4-5 stars
Main result (Table 3)	1.55 (1.50 to 1.60)	1.49 (1.45 to 1.54)
"Yes or no" measure (reviews with any Russian)	1.52 (1.48 to 1.56)	1.46 (1.43 to 1.50)
Context version of the rule	1.55 (1.50 to 1.61)	1.50 (1.45 to 1.55)
Only reviews with a Kazakh-specific letter (141,470)	1.49 (1.43 to 1.54)	1.44 (1.39 to 1.49)
Only texts that occur once (170,846)	1.58 (1.53 to 1.63)	1.53 (1.48 to 1.58)
Threshold 1 per million	1.57 (1.52 to 1.63)	1.52 (1.47 to 1.56)
Threshold 2.5 per million	1.55 (1.50 to 1.60)	1.49 (1.45 to 1.54)
Threshold 5 per million	1.32 (1.28 to 1.36)	1.29 (1.26 to 1.33)
Threshold 10 per million	1.19 (1.16 to 1.23)	1.17 (1.14 to 1.21)
Appstore only (132,417)	1.49 (1.44 to 1.55)	1.43 (1.39 to 1.48)
Market only (29,097)	1.45 (1.22 to 1.70)	1.40 (1.21 to 1.61)
Mapping only (8,492)	1.96 (1.73 to 2.20)	1.92 (1.71 to 2.16)
Bookstore only (4,996)	2.07 (1.31 to 3.04)	1.92 (1.28 to 2.73)

Dissatisfied reviews contain more Russian in every check (Table 4); the smallest ratios are 1.19 (1 against 5 stars) and 1.17 (1-2 against 4-5 stars), both at 10 per million. The "yes or no" measure gives almost the same ratio as the share, in line with Section 6.2. The ratio shrinks at high thresholds because common Kazakh words that are also Russian dictionary forms start to count as Russian, above all *разы* (as in *Алла разы болсын*, "may God be pleased"; 5,085 uses, 98.6% outside 1-2-star reviews); at 10 per million the rule also adds *супер* (2,052 uses, 97.2% outside 1-2-star reviews), a praise word that her boundary could count as Russian, while *пароль* (329 uses, 69.0% in 1-2-star reviews) works the other way. These thresholds therefore partly break the author's boundary, which is why the threshold was set from her twenty words.

6.5 Which Russian words make the difference

Table 5. Russian words per 100 words in 1-2-star and 5-star reviews (each word in one class only; all words of each group counted together, so the last row differs from the mean shares in Table 3; bootstrap intervals from 1,000 resamples of reviews). The first row, a grouping made after a first review, counts вообще, уже, ещё/еще, даже, очень, просто, только, тоже, также, ведь, же, ну, вот, уж, лишь and ни

Type of Russian word	1-2 stars	5 stars	Ratio (95% CI)
Emphatic particles and degree adverbs (16 words, see the note)	0.25	0.08	3.15 (2.82 to 3.53)
Other function words (и, но, на, в, что, это)	0.63	0.39	1.63 (1.51 to 1.78)
Russian stem with Kazakh ending (приложение-ге)	0.75	0.46	1.61 (1.52 to 1.70)
Other Russian content words (приложение, реклама)	2.90	1.75	1.66 (1.60 to 1.72)
All Russian words	4.53	2.68	1.69 (1.64 to 1.75)

The largest ratio is for Russian emphatic particles and degree adverbs (Table 5). Per 10,000 words, *вообще* occurs 6.46 times in 1-2-star reviews against 0.51 times in 5-star reviews, *уже* 5.77 against 1.28, *еще* 4.54 against 1.01 and *даже* 2.16 against 0.59, while the conjunctions *и* (10.39 against 6.54) and *но* (4.69 against 3.59) and the preposition *на* (2.16 against 2.04) differ much less. These words carry no topic of their own and are often used for emphasis (*вообще*, *даже*), although *уже* and *еще* can also mark time ("already", "still, again"); how each was used was not checked. In absolute terms, 1-2-star reviews contain 1.85 more Russian words per 100 words than 5-star reviews: the particles account for 0.17 of them (9%), other function words for 0.24 (13%), mixed words for 0.28 (15%) and other content words for 1.15 (62%), such as *реклама* (10.10 against 1.91 per 10,000 words) and *обновление* (5.12 against 0.73). Russian words also stand somewhat more often next to another Russian word in dissatisfied reviews: 25.0% of the Russian words in 1-2-star reviews are in runs of two or more Russian words, against 20.2% in 5-star reviews (difference 4.8 percentage points, 95% CI 3.1 to 6.7). The Kazakh word of thanks *рахмет* / *рақмет* occurs 129.8 times per 10,000 words in 5-star reviews and 10.5 times in 1-2-star reviews.

7. Discussion

The main result supports the author's hypothesis in direction: the more dissatisfied the writer, the larger the share of Russian words in a Kazakh review. The share falls almost step by step across the ratings and the difference holds in every check of Table 4, but the forecast of three times was too high: the observed ratio is about one and a half. The difference lies mainly in whether a dissatisfied writer uses any Russian at all; once a review mixes, the share of Russian words is only slightly higher, most in long reviews.

The reasons the author gave before the count can now be compared with the data. Her remark that thanks prevail in satisfied Kazakh reviews is clearly supported: *рахмет* is more than twelve times as frequent in 5-star reviews as in 1-2-star reviews, while Russian *спасибо* occurs 0.7 times per 10,000 words in 5-star reviews. Her idea that Russian supplies the words for discontent is supported in part. In absolute terms most of the extra Russian words are content words, many of them technical nouns for the things people complain about (*реклама*, *обновление*, *логин*, *заказ*), which may be Russian because the technology is often named in Russian, not because the writer is angry. But relative to how often they occur, the largest excess

is in Russian emphatic particles (*вообще*, *уже*, *еще*, *даже*), which carry no topic of their own and fit her reason for the hypothesis, that people switch to Russian "to emphasize their dissatisfaction". This study cannot fully separate emotion from topic (Section 8).

The direction connects to Section 2.1: negative opinion moved towards the native language in Hindi-English tweets [5] and contempt towards the less dominant language in a Spanish/Galician conversation [6]; here it moves towards Russian, but the writers' first language is unknown, so the result does not decide between these two pictures.

The result also adds to the author's first study, which found that Russian usually enters a Kazakh sentence as one bare word [1]. In dissatisfied reviews Russian words stand somewhat more often in multi-word Russian stretches (25.0% of Russian words against 20.2%), so single-word insertion remains the usual form of mixing in both groups, with longer Russian stretches a little more frequent when the writer is unhappy.

8. Limitations

First, the automatic rule agreed with a word-by-word reading on 98% of words (90% precision, 83% recall), and that reading was made by the AI assistant that wrote the code, not by a human reader. Second, "established in Kazakh" was measured by Kazakh Wikipedia, which is formal written Kazakh; everyday loanwords that Wikipedia uses rarely, such as *магазин*, are kept as Kazakh only because the threshold was set just below *магазин* (1.85 per million). Third, the star rating measures satisfaction with a product or service rather than emotion as such (Section 4.3). Fourth, the topic of a review is mixed with its rating: complaints are often about technical problems with Russian names. Fifth, three quarters of the reviews are about mobile applications, and nothing is known about the writers. Sixth, KazSAnDRA was collected as reviews in Kazakh [3], so a review written only in Russian is almost invisible here; the true difference may be larger. Seventh, the bootstrap treats reviews as independent, although the dataset paper reports, for the app-store reviews alone, 47,887 named users and 1,759 applications [3]; reviews by the same person or about the same application are alike, so the confidence intervals are too narrow. Finally, this is an observational study: it shows that dissatisfaction and Russian go together, not that dissatisfaction causes the switch.

9. Conclusion

In 175,002 Kazakh online reviews, dissatisfied writers mixed in about one and a half times as much Russian as satisfied writers: Russian words made up an average of 3.85% of the words of a 1-star review against 2.49% of a 5-star review. The author's prediction of the direction was right; her forecast of three times was about twice the observed ratio. The difference comes mainly from how often a review contains any Russian and is larger in long reviews than in short ones; relative to how often they occur, it is largest for Russian emphatic particles such as *вообще* and *уже* (3.15 times), although most of the extra Russian words are content words. A next step would be to separate emotion from topic, for example by comparing reviews of the same application that differ only in rating, or by using a second emotion signal next to the stars.

Data availability. KazSAnDRA is freely available under CC BY 4.0 [3] from its repository [16]. The analysis code, the word lists, the Kazakh Wikipedia frequency list and the per-review counts are kept by the author's research mentor and are available on request.

Author contributions and use of AI tools. Saida Komekova shortlisted two of five research directions prepared by her mentor and agreed to the one he recommended; wrote the hypothesis (18 September 2026) and the reasons for her forecast (4 October), and chose its size, three times, from the two examples in her mentor's question of 5 October (one and a half or

three); chose the share of Russian words over a yes/no measure, the two options her mentor set out, and explained why (25 September; the 100-word and 10-word example came from her mentor's question); drew the boundary between Russian insertions and established loanwords and wrote the two lists of ten words that set the counting rule (27 September; four of her ten Kazakh examples, телефон, магазин, машина and компьютер, were the examples given in her mentor's question of 25 September); found five of the studies cited in Section 2 [6, 7, 9, 10, 11] and wrote a one-line summary of each; wrote the paragraph on star ratings (24 September); summarised the dataset paper (2 October); and approved the analysis plan (6 October). Her paragraphs are quoted in Sections 3, 4.3, 5.2 and 5.3, edited for grammar and spelling, with words added or replaced for meaning in square brackets. Her mentor proposed the five directions, fixed the wording of the question, chose the journal and the three comparison papers from it, and wrote the analysis plan. An AI assistant (Claude, Anthropic), working for her mentor, wrote and ran all of the code (downloading the data, the Kazakh Wikipedia frequency list, applying the counting rule to every word, the recount of the dataset in Section 4.2, the statistics and the figure), drew up the list of Russian function words, grouped the Russian words into the classes of Table 5, made the 46 word corrections and the word-by-word check described in Sections 5.2 and 5.5, found and checked the other references, and drafted the rest of the English text around the author's paragraphs. Apart from the AI assistant's word decisions in Section 5.2 (the 46 corrections and the list of Russian function words), its grouping of 16 words as emphatic particles and degree adverbs in Table 5 and its word judgements in Section 5.5, no number in this paper comes from a language model: every number comes from running the code on the public data, and every reference was opened and checked against its source. K.S. Dyussembina agreed to act as scientific supervisor.

Conflict of interest. The author declares no conflict of interest.

References:

1. Komekova S. Testing classic code-switching constraints on Kazakh-Russian social-media text: an insertional corpus study // Central Asian Scientific Journal. 2026. No. 2(30), Vol. 3. P. 50-58.

2. Smagulova J. Language policies of Kazakhization and their influence on language attitudes and use // International Journal of Bilingual Education and Bilingualism. 2008. Vol. 11, No. 3-4. P. 440-475. DOI: 10.1080/13670050802148798.

3. Yeshpanov R., Varol H.A. KazSAnDRA: Kazakh Sentiment Analysis Dataset of Reviews and Attitudes // Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). Torino: ELRA and ICCL, 2024. P. 9657-9667. URL: https://aclanthology.org/2024.lrec-main.844/ (accessed: 11.10.2026).

4. Dewaele J.-M. The emotional force of swearwords and taboo words in the speech of multilinguals // Journal of Multilingual and Multicultural Development. 2004. Vol. 25, No. 2-3. P. 204-222. DOI: 10.1080/01434630408666529.

5. Rudra K., Rijhwani S., Begum R., Bali K., Choudhury M., Ganguly N. Understanding language preference for expression of opinion and sentiment: what do Hindi-English speakers do on Twitter? // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin: Association for Computational Linguistics, 2016. P. 1131-1141. DOI: 10.18653/v1/D16-1121.

6. Acuña Ferreira A.V. Code-switching and emotions display in Spanish/Galician bilingual conversation // Text & Talk. 2017. Vol. 37, No. 1. P. 47-69. DOI: 10.1515/text-2016-0036.

7. Karimpour S., De Costa P., Nazari M. Emotions in switching the code: contributions for English language teachers' identity construction // Journal of Multilingual and Multicultural Development. 2026. Vol. 47, No. 3. P. 1528-1540 (published online 2024). DOI: 10.1080/01434632.2024.2429413.

8. Gazez A.A. Emotional expression in two languages: bilingual perspectives on language choice and emotional authenticity // Central Asian Scientific Journal. 2026. No. 1(29), Vol. 2. P. 213-217.

9. Lee S.Y.M., Wang Z. Emotion in code-switching texts: corpus construction and analysis // Proceedings of the Eighth SIGHAN Workshop on Chinese Language Processing. Beijing: Association for Computational Linguistics, 2015. P. 91-99. DOI: 10.18653/v1/W15-3116.

10. Tang T., Tang X., Yuan T. Fine-tuning BERT for multi-label sentiment analysis in unbalanced code-switching text // IEEE Access. 2020. Vol. 8. P. 193248-193256. DOI: 10.1109/ACCESS.2020.3030468.

11. Vilares D., Alonso M.A., Gómez-Rodríguez C. EN-ES-CS: an English-Spanish code-switching Twitter corpus for multilingual sentiment analysis // Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). Portorož: ELRA, 2016. P. 4149-4153. URL: https://aclanthology.org/L16-1655/ (accessed: 11.10.2026).

12. Hu N., Zhang J., Pavlou P.A. Overcoming the J-shaped distribution of product reviews // Communications of the ACM. 2009. Vol. 52, No. 10. P. 144-147. DOI: 10.1145/1562764.1562800.

13. Савирдинова А. Сленг в современном языке молодёжи: влияние социальных сетей // Central Asian Scientific Journal. 2026. № 2(30), Т. 1. С. 569-572.

14. Anuarova S.A., Konarova A.B. Language, psyche, and society: cognitive and social interrelations in a multilingual context // Central Asian Scientific Journal. 2026. No. 2(30), Vol. 1. P. 140-151.

15. Assan R. KazSwitch: a controlled benchmark for Kazakh-Russian code-switching in multilingual language models // Central Asian Scientific Journal. 2026. No. 2(30), Vol. 3. P. 65-72.

16. Institute of Smart Systems and Artificial Intelligence (ISSAI). KazSAnDRA: Kazakh Sentiment Analysis Dataset of Reviews and Attitudes. GitHub repository. URL: https://github.com/IS2AI/KazSAnDRA (accessed: 11.10.2026).

17. Korobov M. Morphological analyzer and generator for Russian and Ukrainian languages // Analysis of Images, Social Networks and Texts (AIST 2015). Communications in Computer and Information Science. Cham: Springer, 2015. P. 320-332. DOI: 10.1007/978-3-319-26123-2_31.

18. Wikimedia Foundation. Wikipedia (Kazakh), dump of 01.11.2023, dataset "wikimedia/wikipedia", configuration 20231101.kk. URL: https://huggingface.co/datasets/wikimedia/wikipedia (accessed: 11.10.2026).

UDC 004.8:811.512.122

Zhumagali Yerbol
12th-grade student
Nazarbayev Intellectual School of Science and Mathematics in Turkistan
(Turkistan, Kazakhstan)

DOES A BIGGER MODEL CLOSE THE GAP? KAZAKH-LANGUAGE ACCURACY ON KAZAKHSTAN-SPECIFIC VERSUS UNIVERSAL SCHOOL SUBJECTS IN SMALL OPEN-WEIGHT LANGUAGE MODELS

Abstract: Open-weight language models are usually less accurate in Kazakh than in Russian, but a single Kazakh-versus-Russian score cannot tell whether the model struggles with the language itself or with knowledge about Kazakhstan. We separate the two with a difference-in-differences design on KazMMLU, a bilingual school-level benchmark. Kazakh History, a Kazakhstan-specific subject available in both languages, is compared with mathematics, physics and chemistry, which are Kazakhstan-neutral. We evaluate three sizes of the same open model family (Qwen2.5-Instruct 1.5B, 3B and 7B) with deterministic multiple-choice scoring. On universal subjects the gap between Kazakh and Russian stays within about 1.5 percentage points at every size. On Kazakh History, Kazakh is 13 to 18 points behind Russian, and the gap does not shrink as the model grows: the extra Kazakh penalty on Kazakhstan content is −12.6, −12.9 and −16.7 points (95% CI excludes zero for all three). Scaling within this family improves knowledge about Kazakhstan mainly through Russian. We discuss training-data imbalance as the likely cause and the limits of a single-family, single-subject design.

Keywords: Kazakh language, language models, KazMMLU, low-resource languages, regional knowledge, model scaling

1. Introduction

Language models are now used by students in Kazakhstan for homework, exam preparation and everyday questions, often in a mix of Kazakh and Russian. Models are known to perform worse in Kazakh than in Russian. What is less clear is where that gap comes from. A model may be weaker in Kazakh because it handles the language poorly, or because it knows less about Kazakhstan when the question is asked in Kazakh. The two explanations call for different fixes, yet a single aggregate score mixes them.

KazMMLU was built to measure exactly this kind of knowledge. It contains 23,000 questions from authentic Kazakhstani educational materials, 10,969 in Kazakh and 12,031 in Russian, covering STEM, humanities and social sciences [1]. Its authors report that even strong multilingual models struggle in both languages. Work that uses KazMMLU usually reports a pooled score. For example, Mussa et al. [2] evaluate eleven open-weight checkpoints and use the Kazakh portion of KazMMLU (12 subjects, 9,870 questions) as one measure of general knowledge, reporting Kazakh-only accuracy. Related work on Kazakh and Russian does separate regional from general content, but for safety rather than accuracy (Qorgau [3]). To our knowledge, no study has split the Kazakh-versus-Russian accuracy gap by Kazakhstan-specific versus universal subjects, or asked how that split changes with model size.

We ask: does scaling a small open-weight model close the Kazakh-versus-Russian accuracy gap equally for universal school subjects and for Kazakhstan-specific ones?

2. Data and Method

Benchmark. We use the Kazakh and Russian subsets of KazMMLU. The two subsets are not translations of each other: they have no shared item identifiers and come from different sources. A raw Kazakh-versus-Russian gap therefore mixes the language effect with a difference in item difficulty.

Design. To separate the two, we compare two kinds of subjects, each available in both languages. The Kazakhstan-specific cell is Kazakh History. The universal cell combines Mathematics, Physics and Chemistry, which do not depend on knowledge about Kazakhstan and carry the least language load. The quantity of interest is the difference-in-differences:

DiD = (Kazakh − Russian accuracy | Kazakh History) − (Kazakh − Russian accuracy | universal subjects)

Any item-source difference that is common to all subjects cancels out. What remains is the additional penalty that Kazakh carries on Kazakhstan content. The subject split and the analysis were fixed before the results were seen. Geography, Law and World History mix Kazakhstan-specific and general content and were kept out of the main comparison.

Models. Qwen2.5-Instruct at 1.5B, 3B and 7B parameters [4], 4-bit quantised, run locally. Using one family at three sizes lets us look at scale while holding the training recipe roughly constant.

Scoring. We use standard multiple-choice scoring. Each prompt is written in the item's own language, with three examples from that subset's development split. The model reads the question once, and the predicted answer is the option letter with the highest probability among the letters the item actually has. No text is generated and nothing is sampled, so the run is fully reproducible.

Sample. For each model we score 300 four-option Kazakh History items and 900 universal items (300 per subject) in each language. The main analysis uses four-option items only, because the two languages differ in their share of four- and five-option items and therefore in the chance level. Results on all items are reported as a robustness check. Standard errors follow the binomial formula; the DiD standard error combines the four cells.

Check for answer bias. For each model and language we record the share of answers that go to the most frequently chosen letter (0.25 means no letter preference).

3. Results

Table 1 – Accuracy on four-option items, by model, subject group and language

Model	Kazakh History, Kazakh	Kazakh History, Russian	Gap	Universal gap	DiD (95% CI)
Qwen2.5-1.5B	24.7%	38.7%	−14.0 pp	−1.4 pp	−12.6 pp (−21.1 to −4.0)
Qwen2.5-3B	26.0%	39.3%	−13.3 pp	−0.4 pp	−12.9 pp (−21.6 to −4.2)
Qwen2.5-7B	29.7%	47.3%	−17.7 pp	−1.0 pp	−16.7 pp (−25.6 to −7.7)

Three results stand out. First, on universal subjects the language gap is small at every size (−1.4, −0.4 and −1.0 points). The models read and answer Kazakh school questions about

mathematics, physics and chemistry about as well as Russian ones. Second, on Kazakh History the Kazakh penalty is large, 13 to 18 points, and the difference-in-differences is negative and clearly different from zero for all three models. Third, the penalty does not shrink with scale. From 1.5B to 7B, Kazakh History accuracy rises by 8.6 points in Russian but only 5.0 points in Kazakh, so the largest model shows the largest gap.

The pattern holds on all items, four- and five-option together: DiD −10.3 points ($z = -2.71$), −9.3 points ($z = -2.41$) and −16.1 points ($z = -4.08$).

The answer-bias check shows that the smallest models prefer one letter more often in Kazakh than in Russian (1.5B: 44% vs 31%; 3B: 35% vs 28%), while the 7B model shows no difference (30% vs 30%). Part of the small models' Kazakh gap may therefore reflect guessing behaviour rather than missing knowledge. The 7B result, where this bias disappears, is the cleanest estimate, and it is also the largest.

4. Discussion

Before running the experiment, we expected Kazakh to fall behind everywhere, in history and in the exact sciences alike, and expected the gap to narrow as the model grew, since larger models usually handle rare languages better. Neither expectation held. The gap of about one point in mathematics, physics and chemistry shows that the models understand questions in Kazakh reasonably well; the problem is knowledge of Kazakhstan-specific content. And growing the model did not close that gap: at 7B it was wider than at 1.5B (18 points against 14).

Why would a larger model gain about nine points in Russian on Kazakh History but only about five in Kazakh? We think the answer lies in the training data. Far more text about the history of Kazakhstan exists in Russian than in Kazakh in the corpora such models learn from. A larger model can hold more facts, but it can only remember what it has seen, so in Russian the extra capacity turns into knowledge, while in Kazakh there is little to remember. Kazakh is also tokenised less efficiently than Russian, which may further lower quality. We did not measure the training data or tokenisation directly, so this explanation remains a hypothesis that fits the results rather than a tested claim.

The practical consequence is concrete. A Kazakhstani student who asks a small open model about the history of their own country gets a noticeably worse answer in Kazakh than in Russian, and switching to a bigger model of the same family does not fix it. Making regional knowledge available in Kazakh, rather than only increasing model size, appears to be what matters.

5. Limitations

Sample size. Each cell contains about 300 questions, so every accuracy carries an uncertainty of roughly ±2.5 to 3 points. The comparison between a nine-point and a five-point gain may be partly noise, and the 3B model barely improves over 1.5B in Russian (+0.6 points), so the growth appears mostly at 7B. The main finding, a 13 to 18 point gap on Kazakh History against about one point on universal subjects, is well outside this uncertainty. A repeated run or a larger item sample would sharpen the scale comparison.

One model family. Only Qwen2.5 was tested. Whether the pattern holds for other families remains open.

One Kazakhstan-specific subject. Kazakh History is the only subject that is clearly about Kazakhstan; other mixed subjects were excluded from the main contrast.

Non-parallel subsets. The design removes only the item-source difference that is shared across subjects.

Quantisation. All models were run in 4-bit form, which may lower absolute accuracy.

Answer bias in the smallest models, described above.

6. Conclusion

In three sizes of the same open model family, the gap between Kazakh and Russian accuracy is close to zero for universal school subjects but 13 to 18 points for Kazakh History, and it does not close as the model grows. The weakness is not the Kazakh language as such but knowledge about Kazakhstan expressed in Kazakh. Scaling alone does not solve it.

References:

1. Togmanov M., Mukhituly N., Turmakhan D. [et al.] KazMMLU: Evaluating Language Models on Kazakh, Russian, and Regional Knowledge of Kazakhstan // Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics. Vol. 1. – 2025. – P. 14403–14416. – URL: https://arxiv.org/abs/2502.12829 (accessed: 05.10.2026).

2. Mussa A., Tuimebayev Z., Mansurova M., Shihab A.H. When Does Fine-Tuning Matter for a Low-Resource Language? Evaluating Open-Weight LLMs on Kazakh Benchmarks // Information. – 2026. – Vol. 17, No. 9. – Art. 921. – URL: https://www.mdpi.com/2078-2489/17/9/921 (accessed: 05.10.2026).

3. Goloburda M., Laiyk N., Turmakhan D. [et al.] Qorgau: Evaluating LLM Safety in Kazakh-Russian Bilingual Contexts [Electronic resource]. – 2025. – arXiv:2502.13640. – URL: https://arxiv.org/abs/2502.13640 (accessed: 05.10.2026).

4. Qwen Team. Qwen2.5 Technical Report [Electronic resource]. – 2024. – arXiv:2412.15115. – URL: https://arxiv.org/abs/2412.15115 (accessed: 05.10.2026).

UDC 796:316.77

Abdrakhmanova Madina
Graduate of
Gymnasium No. 17 named after Akan Kurmanov
Scientific supervisor: Tahira Nawaz
(Astana, Kazakhstan)

DOES A MEDAL BUY ATTENTION? WIKIPEDIA VIEWS OF KAZAKHSTAN'S ATHLETES BEFORE AND AFTER THE PARIS 2024 OLYMPIC GAMES

Abstract: This study asks how much public attention an Olympic medal brings to an athlete and how long it lasts, using Kazakhstan's team at the Paris 2024 Olympic Games. Attention was measured as the number of daily views of each athlete's Wikipedia article in Russian, Kazakh and English. The sample covers 58 athletes who competed, 8 of them medalists. Before the data were analysed, I recorded two predictions: strong attention after a medal lasts about two months, and female medalists receive no less attention than male medalists, although this may depend on the sport and on how visible the athlete is in the media. On the peak day, the median medalist received 12,268 views, almost ten times as many as the median of the other athletes (1,275; Mann-Whitney p = 0.001). The peak itself was short: daily views fell to a tenth of the peak within about five days for medalists and four days for the other athletes. In the second month after the medal, however, medalists still received about five times their pre-Games level and in months four to six about three times, and their attention took a median of five months to fall to twice the pre-Games level, compared with about two and a half weeks for the other athletes. The second prediction could not be tested, because only two of the eight medalists were women and their Russian articles did not yet exist during their medal week. The boxer Karina Ibragimova, who did not win a medal, had a higher peak than five of the eight medalists, although several of those medalists were measured without Russian views. A medal increased attention strongly, but the results suggest that it is not the only thing that decides who is noticed.

Keywords: Olympic Games; public attention; Wikipedia page views; athletes; Kazakhstan; sports media; gender

1. Introduction

I have trained in aesthetic group gymnastics for six years and became an Asian champion in 2026. My sport is not part of the Olympic programme, and from my own experience I know that even a high result does not automatically bring attention from the media or from sponsors. In our team, we mostly pay for trips to competitions, uniforms and other expenses ourselves. This made me curious about the opposite case: how much attention does an athlete actually get after winning an Olympic medal, and how long does it last?

Attention is hard to measure directly, but researchers have found a practical proxy. The number of times a Wikipedia article is opened shows how many people actively looked for information about a topic. Views of film articles on Wikipedia helped to predict box office success before release [1], and changes in views of financial articles appeared before stock market moves [2]. Over periods of years, attention to songs, films and biographies decays in two stages, a fast fall followed by a much slower one [3]. For the Paris 2024 Games specifically,

online public attention in China rose before the Games, peaked during them and declined after, closely following the key competition days [4].

Two other lines of research are relevant. In US television sports news, women's sport receives far less coverage than men's sport [5]. On Wikipedia, women are covered in many language editions but are described differently from men [6], and a study of Wikipedia biographies found a persistent gender disparity in the global coverage of Olympic medal winners [7]. For sponsors, an athlete's value is not only the result: a model of athlete brand image includes athletic performance, attractive appearance and a marketable lifestyle [8].

I did not find a study that measured how Olympic medals change public attention to athletes from Kazakhstan. This paper asks three questions. How much more attention does a medal bring? How long does this attention last? And is it the same for women and men?

2. Data and methods

2.1. Athletes

The sample consists of Kazakhstan's athletes at the Paris 2024 Olympic Games who have an article in the English Wikipedia today and are linked from the page «Kazakhstan at the 2024 Summer Olympics» [9]. All links from this page were collected, only people were kept (using Wikidata), and eleven foreign opponents who were linked in the result tables were removed. Two tennis players, Elena Rybakina and Yulia Putintseva, withdrew before their first match [9] and were also removed, because they did not compete. This left 58 athletes: 16 women and 42 men. Kazakhstan won seven medals in Paris: one gold, three silver and three bronze [9]. One bronze was won by a mixed team in shooting, so it is counted for both members, which gives 8 medalists and 50 other athletes.

2.2. Measure of attention

For each athlete, the daily number of views of their article in the Russian, Kazakh and English Wikipedia was taken from the Wikimedia page views service [10], counting only views by people (not automated programs), from 1 May 2024 to 31 July 2025. The three languages were added together. Russian gave most of the views: in August 2024, 65% of all views of medalists' articles were in Russian, 31% in English and 4% in Kazakh. Nineteen athletes had no Russian views in the data and 38 had no Kazakh views, mostly because the article in that language did not exist during this period.

This matters for four medalists. The Russian articles about the shooters Alexandra Le and Islam Satpayev and the boxer Nazym Kyzaibay received their first views only on 12 August, 16 September and 16 September 2024, after their medals, and the silver medalist in gymnastics, Nariman Kurbanov, had no Russian article at all. The medal weeks of Le and Kyzaibay were measured from English only. The peaks of these medalists are therefore underestimated. Because many athletes without a medal also had no Russian article, this problem affects both groups, and its overall direction is not certain.

2.3. Measures

The analysis plan was written on 7 October 2026, before any result was calculated, and my predictions were recorded on 5 October 2026. For each athlete:

• the baseline is the average daily views from 1 June to 20 July 2024, before the Games;

• the peak is the highest daily number of views from the day before the medal to seven days after it; for athletes without a medal, the peak is taken over the whole Games, from 26 July to 11 August 2024;

• the spike is the peak divided by the baseline;

• the afterglow is the average daily views in days 8-30, 31-60, 61-90 and 91-180 after the medal (or after 11 August for other athletes), divided by the baseline;

• the return time is the number of days after the peak until the seven-day average falls to twice the baseline or lower; if this did not happen before the end of the data, the return time is recorded as longer than the observed period.

Because the groups are small and the numbers are very uneven, medians were compared with the Mann-Whitney test, which does not assume a normal distribution. When the baseline was zero, it was set to one view per day to avoid dividing by zero. Several measures were added after the plan was written and are reported as exploratory: the number of days until daily views fall to a tenth of the peak, the comparison of raw peaks between women and men, and the checks described in Section 3.2.

3. Results

3.1. A medal multiplies attention

On their peak day, the median medalist received 12,268 views, compared with 1,275 for the median athlete without a medal, almost ten times as many (Table 1). The difference is statistically significant (Mann-Whitney U = 344, p = 0.001), and the planned spike measure shows the same: 1,408 times the baseline for medalists against 87 times for other athletes ($p < 0.001$). The gold medalist in judo, Yeldos Smetov, had the highest peak of the whole team, 72,575 views in one day. In June 2024 his Russian article was opened about 2,700 times in the whole month; in July, the month of his gold, it was opened more than 130,000 times.

Table 1. Attention to medalists and other athletes, medians

Measure	Medalists (n = 8)	Other athletes (n = 50)	p-value
Peak daily views, median	12,268	1,275	0.001
Spike, multiple of baseline	1,408	87	< 0.001
Days 8-30, multiple of baseline	25.6	1.0	< 0.001
Days 31-60, multiple of baseline	5.2	0.5	< 0.001
Days 61-90, multiple of baseline	4.5	0.5	< 0.001
Days 91-180, multiple of baseline	3.1	0.6	< 0.001
Days until return to 2x baseline	151	17	< 0.001

Note: p-values from the two-sided Mann-Whitney test. Baseline = average daily views from 1 June to 20 July 2024.

Figure 1 shows the whole path of attention. For both groups the peak is sharp: in an exploratory measure, daily views fell to a tenth of the peak in a median of 5.5 days for medalists and 4 days for other athletes. The difference is in what remains after the peak.

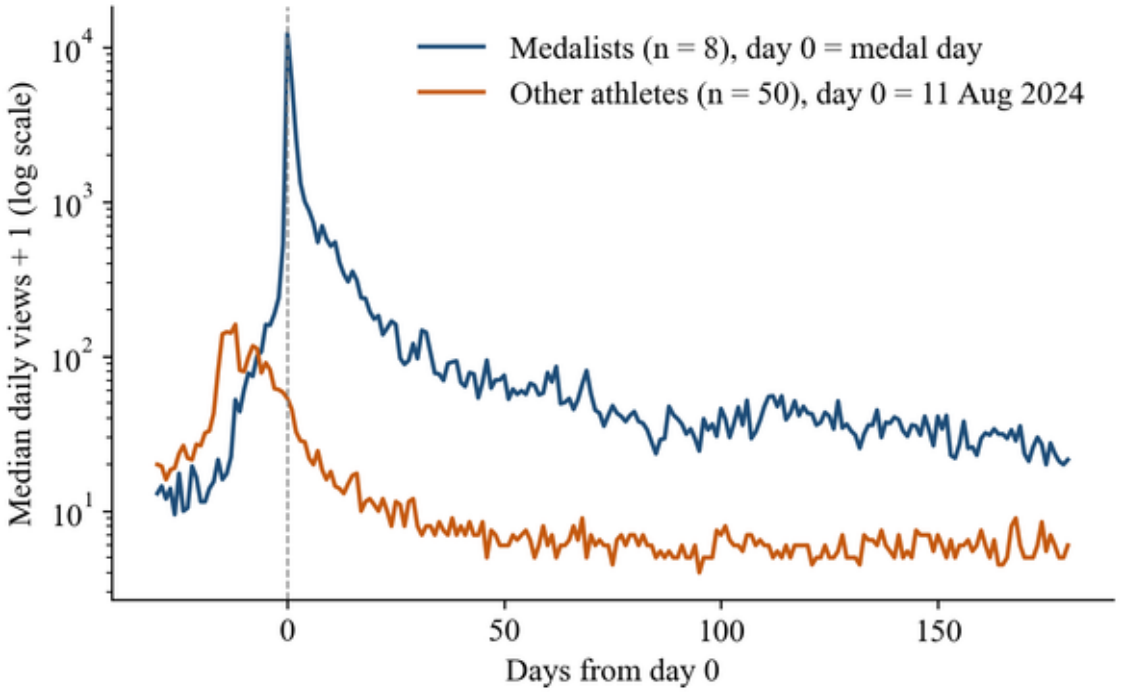


Figure 1. Median daily Wikipedia views + 1 (Russian, Kazakh and English together, log scale).

Day 0 is the medal day for medalists and 11 August 2024, the last day of the Games, for other athletes, so the other athletes' peak during the Games appears before day 0.

3.2. The tail lasts months

In the second month after the medal, the median medalist still received about five times as many views as before the Games, and in months four to six about three times as many. For athletes without a medal, attention was already below the pre-Games level in the second month. The median medalist needed 151 days, about five months, until the seven-day average fell to twice the baseline. The shortest return time was 21 days (Nazym Kyzaibay), and for Alexandra Le this level was not reached in the 369 days of data after her medal. For other athletes the median return time was 17 days.

Some athletes had no views at all before the Games, because their article did not exist yet. The English articles about Alexandra Le and Islam Satpayev received their first views on 27 July 2024, the day of their bronze medal. For nine athletes without a medal the baseline was also zero, and three athletes without a medal (Andrey Chugay, Vadim Sharlaimov and Samirkhon Ababakirov) had no views during the Games, because their articles were created later. Two checks show that this does not change the conclusions. Without all athletes whose baseline was zero (6 medalists and 41 other athletes), the median peak was 17,199 against 1,730 ($p = 0.001$). For the six medalists with a baseline, the afterglow was about 5 times the baseline in the second month and 2.8 times in months four to six, and the return time was 127 days.

Without the three athletes with no views, the other athletes' median peak was 1,491 and their return time 18 days.

3.3. Women and men

In the planned comparison of the spike, women and men did not differ significantly: the median spike was 90 for women and 119 for men (Mann-Whitney $p = 0.96$). In exploratory comparisons, the raw peak also did not differ (1,451 and 1,766; $p = 0.88$), and among athletes without a medal women had a slightly higher median peak than men (1,451 and 1,111; $p = 0.66$). These results are not evidence that there is no difference, only that this sample of 16 women and 42 men does not show one.

My prediction was about female medalists, and it could not be tested. Only two of the eight medalists were women, both bronze medalists. Their peaks, 3,656 views for Alexandra Le and 1,094 for Nazym Kyzaibay, were lower than the median of the six male medalists, 17,199, but for both women the Russian article did not yet exist during the medal week (Section 2.2), so these numbers cannot be compared fairly with the male medalists.

3.4. Attention without a medal

Figure 2 shows the peak of every athlete. The boxer Karina Ibragimova did not win a medal, but her peak, 15,882 views on 2 August 2024, was higher than that of five of the eight medalists. The judoka Abiba Abuzhakynova had a peak of 10,717 views on 27 July 2024, higher than three medalists. That was the day of her own competition, in which she lost the bronze medal contest and finished fifth [9], and also the day Smetov won gold in judo. Both comparisons need care: four of the five medalists that Ibragimova outscored, and all three that Abuzhakynova outscored, were measured without Russian views. Among the four medalists whose Russian article existed during their medal week, Ibragimova had a higher peak than one, and Abuzhakynova than none.

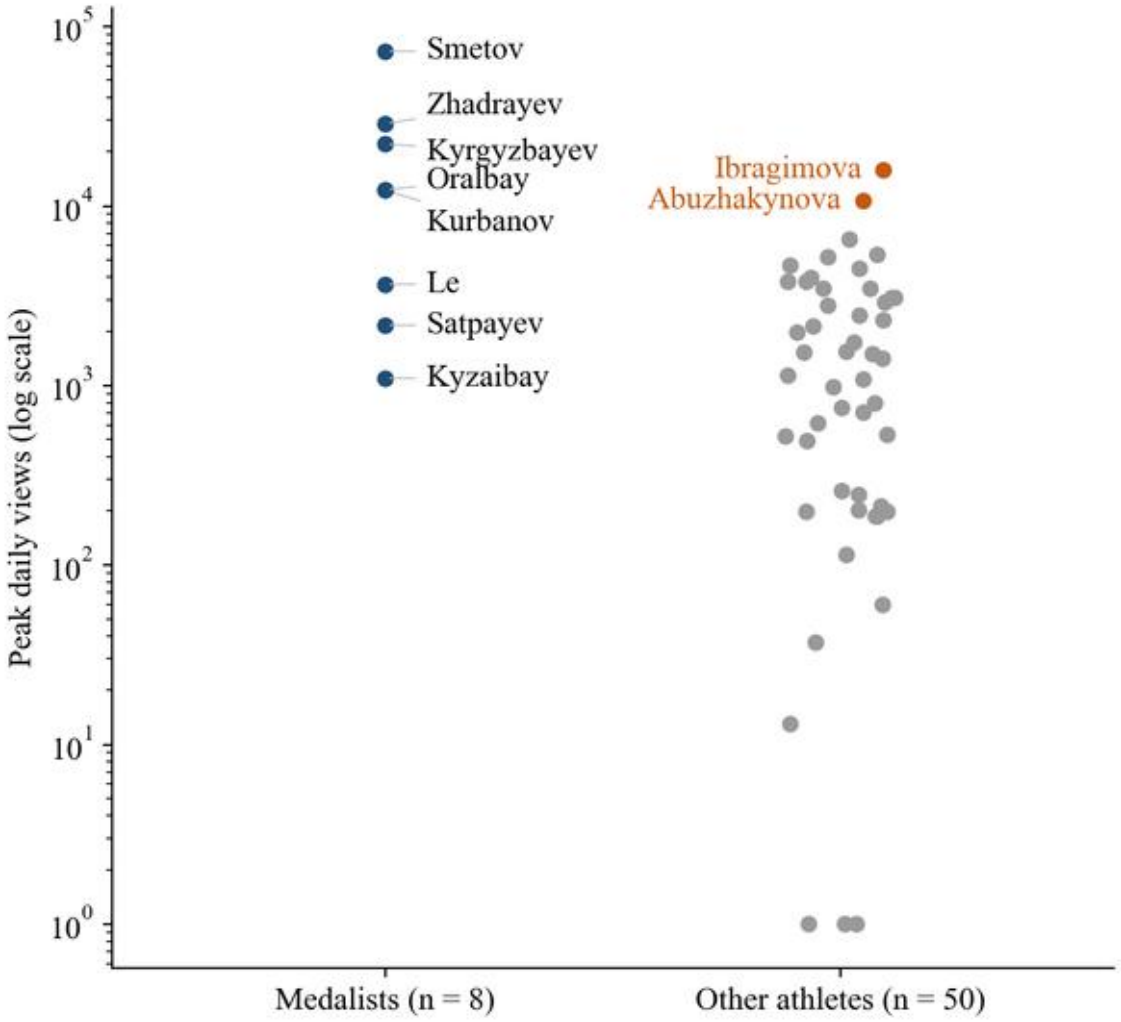


Figure 2. Peak daily Wikipedia views of 58 athletes in their peak window (log scale).

Medalists in dark blue, Ibragimova and Abuzhakynova in orange; athletes with no views are shown at 1.

4. Discussion

My first prediction was that strong attention after a medal would last about two months. The data show a different shape. The wave itself is much shorter, a few days, but the tail is longer than I expected: medalists keep several times more attention than before the Games for about five months. This two-stage pattern, a fast fall and then a slow one, loosely resembles what was found for songs, films and biographies over much longer periods [3]. For an athlete or a sponsor, this means that the first days after a medal are the moment of maximum attention, but a medalist remains noticeably more visible for about half a year.

My second prediction was that female medalists receive no less attention than male medalists. Kazakhstan had only two female medalists in Paris, and for both of them the Russian article appeared only after the medal, so the data cannot answer this question. The Russian articles about both female medalists, and about one male medalist, appeared only weeks after their medals, and one more male medalist had no Russian article at all, while the articles about the other four male medalists already existed. This fits the finding that coverage of female Olympic medal winners on Wikipedia lags behind [7], but with so few people it is an observation, not a result. A study of several Olympic Games would be needed to test it properly.

I also expected that attention would depend on the sport and on the athlete's media presence. The cases of Ibragimova and Abuzhakynova are consistent with this, although two cases cannot measure it. In my opinion, people followed not only their results but also how they competed and how they lived through their defeats. I think many fans expected medals from them, so their bouts were discussed actively. From my experience in gymnastics, spectators sometimes remember not the victories themselves but the emotions of the athletes and how they fought until the end. This agrees with the idea that an athlete's value for the public and for sponsors includes more than athletic performance [8].

My own sport shows the other side. Aesthetic group gymnastics is not an Olympic sport, so it receives little attention from the media and from sponsors, and sponsorship is very rare. From my experience, one medal is not enough. What matters is a long and stable result, and also activity on social media: being able to present yourself well, run your own account and be visible in the media. To attract sponsors, an athlete needs either to be consistently among the best and become a champion, or to actively develop their media presence. The fact that four of Kazakhstan's eight Olympic medalists had no article in Russian, the language most of their readers use, during the week of their medal shows that even at the Olympic level visibility is not automatic.

5. Limitations

Wikipedia views measure only one kind of attention, people actively looking for information, and do not include television, news or social media. The sample includes only athletes with an English Wikipedia article, so the least known athletes are missing, and for many athletes the Russian or Kazakh article did not exist during the Games. For medalists who won early in the Games, days 8-30 partly fall within the Games themselves. The baseline period from 1 June to 20 July 2024 included other major events, such as the Tour de France, which may have raised the baseline of some athletes. The number of medalists is small (8), so the results describe the Kazakhstan team at one Olympic Games and should not be generalised without more data.

6. Conclusion

An Olympic medal brought Kazakhstan's athletes almost ten times as many Wikipedia views at the peak as other team members received. The peak itself lasted only a few days, but medalists kept several times more attention than before the Games for about five months, while for other athletes attention fell to twice the pre-Games level within about two and a half weeks. Women and men on the team did not differ significantly in attention, but with only two female medalists, both measured without their Russian articles, the question of whether female medalists receive less attention could not be tested. Two athletes without a medal outscored some medalists, which suggests that a medal is not the only thing that decides who is noticed.

Data availability

All data are publicly available from the Wikimedia page views service [10] and Wikipedia [9]. The list of athletes, the daily view counts and the script that calculates every number in this paper are available from the author on request.

AI and tools disclosure

I chose the research question, wrote my predictions before seeing any results, and wrote, in Russian, my interpretation of the results and the discussion of my own sport. My mentor collected the data from the Wikimedia service, performed the calculations and prepared the figures. The English text was drafted by my mentor with the help of generative AI tools, based on the results and on my own written answers, and I reviewed and approved it; calling this language editing alone would understate the role of these tools. No data, results or citations were

generated by AI: all numbers come from the Wikimedia page views service, and all cited sources were checked against the originals. I take full responsibility for the final manuscript.

Conflicts of interest

The author declares no conflicts of interest.

References:

1. Mestyán M., Yasseri T., Kertész J. Early prediction of movie box office success based on Wikipedia activity big data // PLoS ONE. 2013. Vol. 8, No. 8. Art. e71226. DOI: 10.1371/journal.pone.0071226.

2. Moat H.S., Curme C., Avakian A., Kenett D.Y., Stanley H.E., Preis T. Quantifying Wikipedia usage patterns before stock market moves // Scientific Reports. 2013. Vol. 3. Art. 1801. DOI: 10.1038/srep01801.

3. Candia C., Jara-Figueroa C., Rodriguez-Sickert C., Barabási A.-L., Hidalgo C.A. The universal decay of collective memory and attention // Nature Human Behaviour. 2019. Vol. 3, No. 1. P. 82-91. DOI: 10.1038/s41562-018-0474-5.

4. Zhang M., Liao Y. Social media impact on public attention: Paris Olympics case study // PLoS ONE. 2025. Vol. 20, No. 11. Art. e0336136. DOI: 10.1371/journal.pone.0336136.

5. Cooky C., Messner M.A., Musto M. «It's dude time!»: a quarter century of excluding women's sports in televised news and highlight shows // Communication & Sport. 2015. Vol. 3, No. 3. P. 261-287. DOI: 10.1177/2167479515588761.

6. Wagner C., Garcia D., Jadidi M., Strohmaier M. It's a man's Wikipedia? Assessing gender inequality in an online encyclopedia // Proceedings of the International AAAI Conference on Web and Social Media. 2015. Vol. 9, No. 1. P. 454-463. DOI: 10.1609/icwsm.v9i1.14628.

7. Yu Y., Li X., Li T., Dhillon P.S., Romero D.M. Demographic disparity in Wikipedia coverage: a global perspective // EPJ Data Science. 2025. Vol. 14. Art. 15. DOI: 10.1140/epjds/s13688-025-00530-4.

8. Arai A., Ko Y.J., Ross S. Branding athletes: exploration and conceptualization of athlete brand image // Sport Management Review. 2014. Vol. 17, No. 2. P. 97-106. DOI: 10.1016/j.smr.2013.04.003.

9. Wikipedia. Kazakhstan at the 2024 Summer Olympics [Electronic resource]. URL: https://en.wikipedia.org/wiki/Kazakhstan_at_the_2024_Summer_Olympics (accessed: 13.10.2026).

10. Wikimedia Foundation. Analytics API: page views [Electronic resource]. URL: https://doc.wikimedia.org/generated-data-platform/aqs/analytics-api/reference/page-views.html (accessed: 13.10.2026).

ПЕДАГОГИКАЛЫҚ ЖӘНЕ ПСИХОЛОГИЯЛЫҚ ҒЫЛЫМДАР – ПЕДАГОГИЧЕСКИЕ И ПСИХОЛОГИЧЕСКИЕ НАУКИ – PEDAGOGICAL AND PSYCHOLOGICAL SCIENCES

UDC 37.014:316.343

Kaldarbek Mergen

Grade 11 student
Miras International School Almaty
(Almaty, Kazakhstan)

FAMILY OR SCHOOL? SOCIOECONOMIC COMPOSITION AND MATHEMATICS ACHIEVEMENT IN KAZAKHSTAN, PISA 2022 AND 2025

Abstract: It is widely assumed in Kazakhstan that a family's resources decide a child's school results and that the school itself is a supplement. This study tests that assumption directly on the two most recent PISA cycles. Using the OECD Public Use Files for Kazakhstan, 19,769 students in 571 schools in 2022 and 22,680 students in 643 schools in 2025, the mathematics score is decomposed into the part explained by a student's own family background and the part explained by the average background of the school they attend. The design, the hypothesis, the falsifier and every robustness check were written down and committed before any student-level number was computed. Estimation follows the PISA methodology: ten plausible values combined by Rubin's rules and sampling variance from eighty Fay replicate weights. The pipeline reproduces the OECD's published country means to within half a point in both cycles. The result contradicts the stated assumption and does so by a wide margin. Within a school, one unit of family socioeconomic status is worth 6.4 score points in 2022 and 7.0 in 2025. The school's own mean socioeconomic status is worth 56.4 and 58.1 points, roughly nine times more, and the gap is stable across three years. It survives reading and science, a leave-one-out construction of the school mean, restriction to schools with at least ten sampled students, and controls for school type and community size. Thirty to thirty-four per cent of the variance in mathematics lies between schools. In a regional comparison Kazakhstan has one of the lowest within-school family gradients measured, 7.0 against 24.9 in Estonia, while its contextual effect is among the highest. The practical reading is narrow and should not be overstated: this is an observational association, and the school mean carries peer composition, unobserved school quality and family sorting together. What the data do establish is that in Kazakhstan the question "which school" separates pupils far more sharply than the question "which family", and any policy aimed at the second while ignoring the first is aimed at the smaller of the two.

Keywords: PISA; Kazakhstan; socioeconomic status; school composition; contextual effect; educational inequality; mathematics achievement

Introduction. Ask most people in Kazakhstan why one pupil scores better than another and the answer arrives quickly: the family. A household with money buys tutors, books, a quiet room, a language environment; a household without them cannot, and the school is treated as a background condition that a determined family can compensate for. The assumption is so ordinary that it is rarely stated, let alone tested on national data.

This paper states it and tests it. The author wrote the assumption down as a hypothesis before seeing any result, in these words: a pupil's result in Kazakhstan depends more on the family's resources than on the particular school, because a family with income can compensate for any gap through courses and mentoring, while the school only sets the academic bar.

The question that makes this testable is not whether family background matters. It plainly does, and a large literature has measured how much [1]. The question is whether the socioeconomic character of the school a pupil attends predicts their score **beyond** their own family's position. That quantity has a name in the literature, the contextual or compositional effect, and it has been studied across the PISA countries [2, 3, 4], but not, as far as we could find, reported for Kazakhstan as a separate within-and-between decomposition on the most recent cycles.

It matters which of the two is larger, because the two point at different policies. If the family effect dominates, the lever is the household: targeted allowances, tutoring vouchers, subsidised materials. If the school effect dominates, the lever is the composition of schools themselves, and money handed to a family changes little unless it changes which school the child sits in.

Data. Both cycles come from the OECD PISA Public Use Files: the student questionnaire file, which also carries the performance estimates, and the school questionnaire file. Kazakhstan is identified by the country code `KAZ`. The 2022 files are openly downloadable; access to the 2025 files was granted through the OECD's data-access form, and the raw files are not redistributed with this paper.

The 2022 Kazakhstan sample is 19,769 students in 571 schools, of whom 19,683 carry a socioeconomic index. The 2025 sample is 22,680 students in 643 schools, of whom 21,198 carry the index.

Achievement is measured by the ten plausible values for mathematics, with reading and science used as robustness checks. Socioeconomic status is the OECD's index of economic, social and cultural status (ESCS). The school is identified by `CNTSCHID`, and the school's socioeconomic position is the weighted mean ESCS of its sampled students. All estimates use the final student weight and the eighty Fay replicate weights.

One check was run before anything else. The pipeline was asked to reproduce the country means the OECD has already published. It returned 425.4 in mathematics for 2022 against the published 425, and 413.7 for 2025 against the published 414, with reading and science equally close. A pipeline that cannot reproduce a known number should not be trusted with an unknown one.

I honestly thought that a family's own financial capital decided almost everything, because that is what pays for tutors, equipment and basic freedom. The results showed something quite different: the effect of the school itself turned out to be incomparably stronger than family wealth. What surprised me most is that a school can give that kind of enormous boost regardless of a pupil's individual background, when it seemed that tutors and resources at home ought to outweigh it. Just as unexpected was that private status in itself is no guarantee of an outstanding result, and that what decides it is the environment and the mix of pupils you end up among. This is a very interesting error in my hypothesis, and it shows plainly how strong the systemic influence of the right learning environment is.

The author's own school illustrates why the two quantities are not the same thing. It is a private international school in Almaty; in the IB grade-11 cohort there are 31 pupils; roughly half hold academic grants or fee discounts and the other half pay in full, and for about five or

six families the fee sits at the edge of what they can manage. Family circumstances inside that one room therefore vary widely, while the school itself is fixed. That is exactly the contrast the models below separate.

Method. The pre-registered plan, written and committed before any Kazakhstan student-level number was computed, fixed the hypothesis, the condition that would refute it, the models, the robustness checks and the stated limits. No deviation was made; one amendment, also before any result, added PISA 2025 as the primary cycle once access was granted, leaving every model unchanged.

Every model is weighted least squares with the final student weight. Sampling variance comes from the eighty Fay replicate weights with the standard adjustment, and each model is run on all ten plausible values and combined by Rubin's rules, so the reported standard errors carry both the sampling and the imputation component.

Four models were specified in advance.

M0 splits the variance in mathematics into a within-school and a between-school part.

M1 regresses the score on the pupil's own ESCS. This is the total gradient, and it mixes the two effects that the next model separates.

M2 is the Mundlak decomposition [5]. The score is regressed on the pupil's ESCS measured as a deviation from their school's mean, and on the school mean itself. The first coefficient is the within-school family effect: what one unit of family advantage buys among pupils sitting in the same school. The difference between the two coefficients is the contextual effect: what the school's own composition adds on top. The hypothesis is tested by comparing these two.

M3 repeats M2 with the school's type, public or private, and the size of its community, to see whether the contextual effect is simply those two things wearing a disguise.

Robustness was fixed in advance as well: repeat on reading and science; recompute each school's mean leaving the pupil themself out, so that no one can inflate their own school's average; restrict to schools with at least ten sampled pupils; and drop the community-size category "a megacity of over ten million", which cannot describe any settlement in Kazakhstan and which twenty-one principals nevertheless selected in each cycle.

Results

How much lies between schools

Table 1 — Kazakhstan samples, the reproduction check against the OECD's published means, and the share of achievement variance lying between schools rather than within them.

Quantity	PISA 2022	PISA 2025
Students (with ESCS)	19,683	21,198
Schools	571	643
Mean mathematics score	425.4	413.7
Published OECD mean	425	414
Between-school share of variance, mathematics	30.0%	33.6%
Between-school share, reading	33.1%	34.9%

Between-school share, science	29.4%	33.3%
Top-minus-bottom ESCS quarter gap, mathematics	43.3	44.4

About a third of all variation in Kazakhstani mathematics scores is variation between schools rather than between pupils inside them, and that share rose slightly between the two cycles.

Family or school

Table 2 — Score points per unit of ESCS. The within-school effect is what family background buys among pupils in the same school; the contextual effect is what the school's own average background adds beyond that. Every row uses ten plausible values and eighty replicate weights.

Model	Cycle	Within-school family effect	Contextual school effect
M2 — mathematics	2022	6.37 (SE 0.87)	56.43 (SE 3.99)
M2 — mathematics	2025	7.03 (SE 0.82)	58.13 (SE 4.34)
M3 — + school type and community size	2022	6.47 (SE 0.95)	58.38 (SE 5.44)
M3 — + school type and community size	2025	7.05 (SE 0.86)	60.44 (SE 4.20)
Reading	2022	7.55 (SE 1.11)	62.72 (SE 4.78)
Reading	2025	5.13 (SE 0.84)	48.86 (SE 3.57)
Science	2022	7.02 (SE 1.02)	56.15 (SE 3.99)
Science	2025	7.50 (SE 0.83)	52.08 (SE 3.79)
Leave-one-out school mean	2022	8.39 (SE 0.86)	54.14 (SE 3.86)
Leave-one-out school mean	2025	9.19 (SE 0.79)	55.59 (SE 4.19)
Schools with 10+ sampled students	2022	6.19 (SE 0.89)	60.00 (SE 4.20)
Schools with 10+ sampled students	2025	7.07 (SE 0.81)	61.07 (SE 4.47)

The hypothesis is refuted, and not narrowly. Among pupils sitting in the same school, one unit of family advantage is worth about six or seven score points. The school's own average background is worth between forty-nine and sixty-one. The ratio is roughly nine to one, it is the same size in both cycles three years apart, and no robustness check moves it.

The total gradient in M1, 18.73 points in 2022 and 19.29 in 2025, is therefore a blend and a misleading one on its own. Reported alone it suggests a moderate family effect; decomposed, most of it turns out to be the school.

M3 rules out the simplest alternative explanation. Adding whether the school is private and how large its settlement is does not shrink the contextual effect at all; it grows slightly. Private schools score about seventeen to twenty points lower once composition is held constant, which is not a claim about their quality but a reminder that in this sample the private sector is not a uniformly advantaged one.

The leave-one-out check is the one that matters most for honesty, because a pupil contributes to their own school's mean and could mechanically inflate it. Removing them raises the within-school effect by about two points and lowers the contextual effect by about two. The conclusion is untouched.

Kazakhstan against the region

Table 3 — The same decomposition, PISA 2025 mathematics, for the other post-Soviet systems in the file. Uzbekistan is absent: its 2025 design produced a degenerate estimate and was dropped rather than reported.

System	Within-school family effect	Contextual school effect
Kazakhstan	7.0	58.1
Estonia	24.9	49.2
Moldova	15.9	45.6
Mongolia	12.6	51.6
Latvia	12.8	55.0
Lithuania	11.0	83.1
Georgia	10.3	47.1
Armenia	8.5	40.8
Kyrgyzstan	7.5	57.4
Azerbaijan	7.3	77.0

Kazakhstan sits at one extreme of this table. Its within-school family effect, 7.0 points, is among the lowest measured; Estonia's is three and a half times larger. Its contextual effect is in the upper half. The combination is what matters: in Estonia a family's position still moves a pupil noticeably among their own classmates, while in Kazakhstan it barely does, and almost the whole socioeconomic signal has migrated to the level of which school a child attends.

Discussion. The finding has a plain reading and a careful one, and both should be stated.

The plain reading is that in Kazakhstan the sentence "rich families do better" is a poor description of the data, while "children in rich schools do better" is a good one. Once a pupil is inside a given school, how much their own family has adds surprisingly little. Which school they are inside adds a great deal.

The careful reading is that the contextual effect is an association, not a mechanism, and it bundles at least three things that this design cannot separate: genuine peer effects, school quality that is correlated with composition but not caused by it, and the sorting of families into schools by resources and ambition. A wealthy family that moves its child into a high-composition school buys some mixture of all three, and this paper cannot say in what proportion.

That caveat does not dissolve the policy point, it sharpens it. Whatever the mechanism, the measured separation between Kazakhstani pupils runs along school lines far more than

household lines. An intervention that raises household resources without changing which school a child attends is acting on the six-point channel. The nine-times-larger channel lies elsewhere.

The regional comparison suggests this is not a universal pattern but a Kazakhstani one, or at least one shared with a few neighbours. Estonia's profile, a substantial within-school gradient and a moderate contextual effect, looks like a system where schools are more alike and families differentiate pupils inside them. Kazakhstan's looks like the opposite.

Limitations. The design is cross-sectional and observational. Nothing here identifies a causal effect of schools, and the contextual coefficient should not be read as what would happen if a pupil were moved.

The school mean is estimated from the sampled pupils, not the whole school, so it is measured with error. The leave-one-out check addresses one form of this, not all of it.

ESCS is a constructed index combining parental occupation, education and household possessions. It is comparable across the PISA countries by design, but it is not a direct measure of income, and its components may function differently in Kazakhstan than in the OECD average.

The two cycles are independent samples of fifteen-year-olds, not the same pupils followed over time, so the stability reported here is stability of a population pattern, not of individuals.

Finally, the comparison in Table 3 uses the systems present in the 2025 file. Uzbekistan was dropped because its design produced a non-finite estimate; it is named here rather than quietly omitted.

Conclusion. Across the two most recent PISA cycles, the socioeconomic composition of a Kazakhstani pupil's school predicts their mathematics score about nine times more strongly than their own family's socioeconomic position does, and the ratio is stable across three years, three subjects and every robustness check specified in advance. The assumption the study set out to test, that the family decides and the school supplements, is the reverse of what the data show. Roughly a third of all achievement variance in Kazakhstan lies between schools rather than within them, and in a regional comparison the country has one of the weakest within-school family gradients measured alongside one of the stronger contextual effects. The mechanism behind that contextual effect, how much is peers, how much is school quality and how much is families sorting themselves, is the obvious next question and is not answered here.

Data availability and replication

The analysis code (`analysis.py`), the pre-registered plan with its amendment, and the complete machine-readable results for both cycles accompany this paper. The PISA Public Use Files are not redistributed: the 2022 files are openly downloadable from the OECD and the 2025 files are obtained through the OECD's data-access form, in accordance with the terms of use accepted there. Running the code against those files reproduces every number in every table, including the reproduction check against the OECD's published country means.

Declarations

Funding: none. Conflicts of interest: none. Ethics: the study uses only anonymised aggregate survey data published for research; no personal data were collected and no human subjects were involved.

Use of AI tools.The author used generative AI tools to help write the analysis code, run the statistical computations and edit the English text. The research question, the hypothesis recorded before the data were seen, the school description in the author's own account and the interpretation are the author's. Every number reported here is produced by the released analysis

script from the OECD's own files; none was generated by a language model. The OECD is acknowledged as the source of the PISA data.

References:

1. Sirin S. R. Socioeconomic status and academic achievement: a meta-analytic review of research // Review of Educational Research. 2005. Vol. 75, No. 3. P. 417–453. DOI: 10.3102/00346543075003417.

2. Willms J. D. School composition and contextual effects on student outcomes // Teachers College Record. 2010. Vol. 112, No. 4. P. 1008–1037. DOI: 10.1177/016146811011200408.

3. Perry L. B., McConney A. Does the SES of the school matter? An examination of socioeconomic status and student achievement using PISA 2003 // Teachers College Record. 2010. Vol. 112, No. 4. P. 1137–1162. DOI: 10.1177/016146811011200401.

4. Perry L. B., McConney A. School socio-economic composition and student outcomes in Australia: implications for educational policy // Australian Journal of Education. 2010. Vol. 54, No. 1. P. 72–85. DOI: 10.1177/000494411005400106.

5. Mundlak Y. On the pooling of time series and cross section data // Econometrica. 1978. Vol. 46, No. 1. P. 69–85. DOI: 10.2307/1913646.

6. OECD. PISA 2022 Database [Electronic resource]. Paris: OECD, 2023. URL: https://www.oecd.org/en/data/datasets/pisa-2022-database.html (accessed: 25.09.2026).

7. OECD. PISA 2025 Database [Electronic resource]. Paris: OECD, 2026. URL: https://www.oecd.org/en/data/datasets/pisa-2025-database.html (accessed: 25.09.2026).

8. OECD. PISA 2022 Results (Volume I and II), Country Note: Kazakhstan [Electronic resource]. Paris: OECD, 2023. URL: https://gpseducation.oecd.org/ (accessed: 20.09.2026).

UDC 51:37.091.33(574)

Zhenis Adilzhan
Grade 11 student
Supervisor: Kaldanov Serikzhan
mathematics teacher
Bilim-Innovation Lyceum No. 1
(Karagandy, Kazakhstan)

WHICH OLYMPIAD PROBLEMS DO STUDENTS TAKE TO THE BOARD? CHOICES AND DEFENCES IN THE 8TH-GRADE LEAGUE OF KARAGANDY MATHBATTLES 2026

Abstract: A mathematical battle is a team competition in which two teams solve the same set of problems and then challenge each other to present solutions at the board. A challenged team can accept the challenge or refuse it, so every battle records a decision about each problem. In this study I used the jury protocols of the 8th-grade league of Karagandy Mathbattles 2026, a tournament where I was one of the organisers and a member of the jury, to find out which problems teams take to the board and how successful their defences are. I transcribed all 16 battles of the league (8 teams, 96 challenges on 30 different problems), checked every battle against the jury's total, and labelled each problem as algebra, geometry, combinatorics or number theory. Teams accepted 68% of challenges on combinatorics problems, but only 46% on geometry, 42% on algebra and 38% on number theory (combinatorics against the other topics, Fisher's exact test $p = 0.05$). When a geometry challenge was accepted, the speaker almost always gave a complete solution (10 of 11 reports received the full 12 points, mean 10.9), while accepted number theory problems received a mean of 6.4 points and only 3 of 10 full solutions (permutation test $p = 0.03$). In 30 of the 49 refused challenges the challenging team could not defend its own challenge, so the refusing team received 6 points. Stronger teams accepted challenges more often than weaker ones. The results suggest that teams take geometry to the board only when they are sure, take combinatorics more boldly, and avoid number theory, which was also the topic in which students were least successful at the board. Because the data come from one league of one tournament, these patterns should be checked on more tournaments.

Keywords: mathematical battle; mathematics olympiad; problem solving; team competition; problem difficulty; Kazakhstan

Introduction. Most mathematics olympiads are individual: a student writes solutions on paper and the jury marks them later. Competitions have long been an important part of mathematics education [1], and over the last decades many team competitions have appeared next to individual olympiads [2]. One of them is the mathematical battle (in Russian "математический бой", or "matboi"), which comes from the tradition of mathematical circles in the former Soviet Union [3]. In a battle, two teams get the same problems and solve them in separate rooms. After that they meet, and the teams take turns challenging each other: one team names a problem, and the other team decides whether it will present the solution at the board [4, 5].

I have taken part in the organisation of Karagandy Mathbattles, a tournament for students of the 7th and 8th grades, and in 2026 I was also a member of the jury. From the jury table it is

easy to see that teams do not treat all problems in the same way. Some problems are accepted almost every time, while other problems are refused even by strong teams. Every refusal is a decision with a cost, because a team that refuses a problem cannot score the full 12 points for it. This means that the protocols of a battle show not only which problems were solved, but also which problems students felt confident enough to defend in front of an opponent and a jury. In research on problem solving, the ability to judge and control one's own work is called metacognition, and it is considered an important part of thinking mathematically [6].

Olympiad problems are usually divided into four areas: algebra, combinatorics, geometry and number theory. This division is used, for example, in the shortlist of the International Mathematical Olympiad [7]. I did not find a published study that uses the protocols of mathematical battles to compare these areas, and the rules of a battle make such data unusual, because they record both the decision about a problem and the result of its defence. In this study I asked three questions:

• Which topics do teams accept to present at the board, and which do they refuse more often?

• When a challenge is accepted, how successful is the defence in each topic?

• What happens when a challenge is refused, and do stronger teams accept challenges more often?

The rules of a mathematical battle. The rules used at Karagandy Mathbattles follow the usual rules of mathematical battles [4, 5]. Each battle has six problems (in the final the teams had 1 hour 30 minutes to solve them). Each problem is worth 12 points, which are divided between the speaker, the opponent and the jury. A speaker who gives a complete and correct solution receives all 12 points. The opponent's task is to find gaps in the solution, and the opponent receives points for the gaps that he finds. The points that neither side earned stay with the jury.

When a team is challenged, it can accept or refuse. If it accepts, it sends a speaker to the board and the challenging team sends an opponent. If it refuses, the challenging team must show that it has a solution itself. This is called the check of the correctness of the challenge. If the challenging team admits that it has no solution, or if the opponent proves that the solution is not there, the challenge is called incorrect, the refusing team receives 6 points and the challenging team has to make another challenge [5]. The jury can also give penalty points, usually -1, for breaking the rules, for example for talking to a teammate at the board.

Data and methods

The tournament. I used the 8th-grade league of Karagandy Mathbattles 2026. Eight teams took part. The teams played three group rounds of four battles each, then two semifinals, a match for third place and the final, which makes 16 battles. Each battle had six problems, so the data contain 96 challenges. The three group rounds and the semifinals had their own sets of problems, and the final and the match for third place used the same set, so there were 30 different problems in total.

Transcription. After every battle the jury writes the protocol on the whiteboard: for each problem, the number, the names of the speaker and the opponent, the points of each side, the jury's points, penalties and whether the challenge was accepted. These protocols were photographed. I transcribed them into a table with one row per problem in each battle. During this work I found that Google Sheets had changed some entries such as "12-1" (12 points and a penalty of 1) into dates, so I moved penalties into separate columns and entered all points as numbers.

Checks. The sum of the points in the six rows of a battle must be equal to the total written by the jury. After all corrections, this was true in 15 of the 16 battles. In one battle of the first round the protocol said 29 points for one team, but the points for its problems add up to 30, so the jury made an arithmetic error. For the final there is no photo of the protocol, and the score of 31:27 comes from a member of the jury. I kept both of these facts in the data and mention them here. In one challenge, in the match for third place, both teams refused to present the problem. This challenge is counted in the number of challenges but not in the analysis of defences.

Topics. I labelled each problem by its main topic, using the statement and the official solution. Five of the 30 problems combine two topics, for example a problem about subsets that is solved with algebra. For them I chose the main topic by the method of the solution, and as a robustness check I repeated the calculations giving half of such a problem to each of its two topics. By main topic, the 96 challenges include 26 on number theory, 24 on algebra, 24 on geometry and 22 on combinatorics.

Measures and tests. For each topic I calculated the share of challenges that were accepted and, for accepted challenges, the points of the speaker. For refused challenges I counted how often the challenge was found incorrect, which I defined as the refusing team receiving 6 points while the challenging team's speaker received no more than 5. To test whether acceptance depends on the topic I used the chi-square test and Fisher's exact test. Because stronger teams may accept more problems, I also used a permutation test [8] in which the topic labels were shuffled 10,000 times only within the challenges received by the same team. To compare speaker points between topics I used the Kruskal-Wallis test [9] and a permutation test for the difference in means. To describe the strength of a team I used its average number of points per battle. All calculations were made in Python.

Results

Which problems are taken to the board. Teams accepted 46 of the 96 challenges (48%). The share was very different between topics (Table 1 and Figure 1). Combinatorics was accepted in 15 of 22 challenges (68%), geometry in 11 of 24 (46%), algebra in 10 of 24 (42%) and number theory in 10 of 26 (38%). The chi-square test for all four topics was not significant ($p = 0.17$), and the result was the same when topics were shuffled only within each team ($p = 0.17$). However, combinatorics stands out from the other three topics: the odds of accepting a combinatorics challenge were 2.9 times higher than for the other topics together (Fisher's exact test $p = 0.05$). With only 96 challenges, I think this should be read as a clear tendency rather than a proven difference.

Table 1. Challenges by the main topic of the problem

Topic	Challenges	Accepted, %	Mean speaker points (accepted)	Full solutions (accepted)	Incorrect challenges (refused)
Geometry	24	46	10.9	10 of 11	7 of 12
Combinatorics	22	68	8.9	9 of 15	4 of 7
Algebra	24	42	7.9	5 of 10	9 of 14
Number theory	26	38	6.4	3 of 10	10 of 16

One geometry challenge, in which both teams refused, is counted among the challenges but not among accepted or refused ones.

Some problems were never accepted. In the group stage, where each problem was given in four battles, two problems were refused every time: a geometry problem about a rectangle in the first round and an algebraic inequality in the third round. Both number theory problems of the final set were refused in the final and in the match for third place, and the number theory problem of the semifinals was refused in both semifinals.

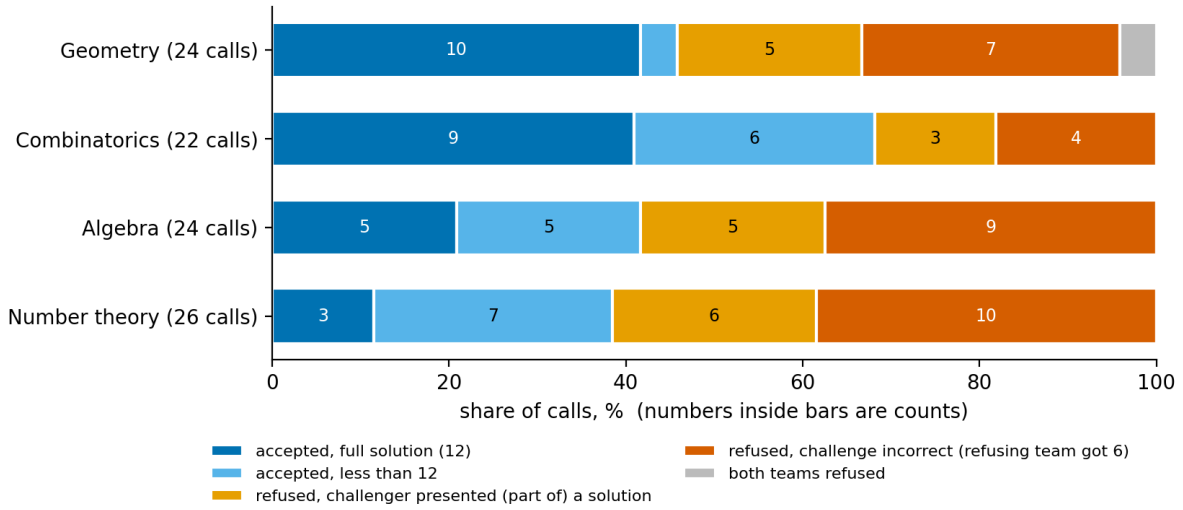


Figure 1. What happened to the challenges in each topic

How successful the defences were. When a challenge was accepted, the speaker received on average 8.6 points, and 27 of the 46 reports (59%) were complete solutions worth 12 points. The topics differed a lot (Figure 2). In geometry 10 of 11 accepted reports received the full 12 points, and the mean was 10.9. In combinatorics the mean was 8.9 with 9 complete solutions out of 15, in algebra 7.9 with 5 out of 10, and in number theory only 6.4 with 3 out of 10. The Kruskal-Wallis test for all four topics gave $p = 0.08$, and the difference between geometry and number theory, 4.5 points, was significant in a permutation test ($p = 0.03$).

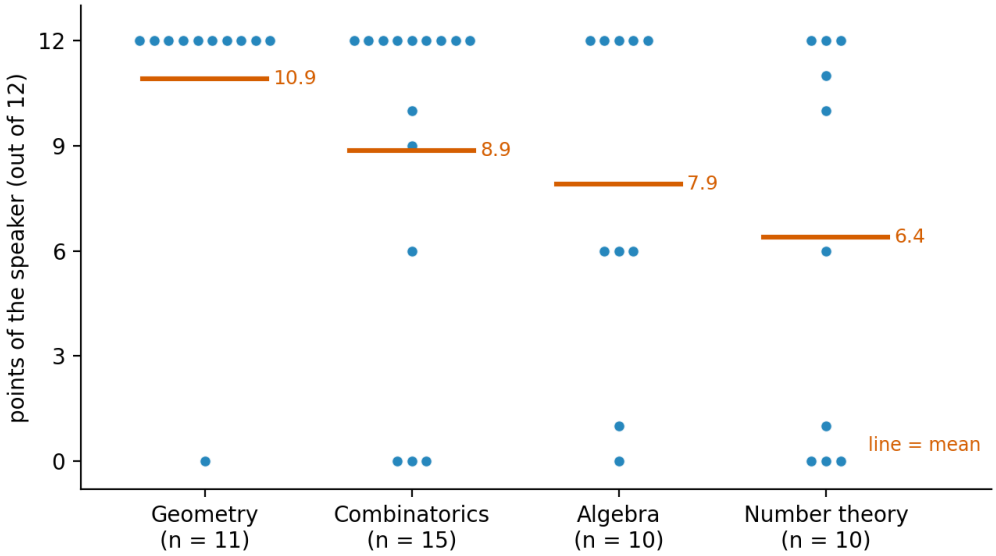


Figure 2. Points of the speaker in accepted challenges, by topic

Geometry and combinatorics therefore behave differently. Geometry was accepted less often than combinatorics, but when it was accepted, the solution was almost always complete. Combinatorics was accepted more often, and more of these defences ended with a partial solution. My explanation is that in geometry students know quite well whether their proof is finished, while in combinatorics it is easier to believe that a solution is correct when it still has a gap, and the opponent can find it. Number theory was both the least accepted topic and the least successful one at the board.

The robustness check with half weights for the five problems with two topics gave the same order: geometry still had the highest mean (10.9) and number theory the lowest (6.6), and combinatorics was still accepted most often (61%).

Refusals and the strength of teams. Of the 49 refused challenges, 30 (61%) turned out to be incorrect: the challenging team did not have a solution, and the refusing team received 6 points. In only 11 refusals did the challenging team present a complete solution itself. Incorrect challenges were most common in number theory (10 of 16) and algebra (9 of 14). This means that often neither team had a solution to the problem, and the refusal was a reasonable decision. On average, a challenged team received 8.6 points when it accepted and 3.8 points when it refused.

Stronger teams accepted challenges more often (Table 2). The three medal teams accepted between 47% and 61% of the challenges they received, while KarErBIL-2 and Daryn, two of the weakest teams, accepted 25% and 33%. The weakest team, Burabay, is an exception with 50%, but it received only 8 challenges. The rank correlation between a team's average points per battle and its acceptance rate was 0.63, but with only eight teams this was not significant ($p = 0.09$). Because strong teams play more battles, I also checked the topic results within teams, as described in section 3, and the conclusions did not change.

Table 2. Teams of the 8th-grade league

Team	Battles	Points per battle	Challenges received	Accepted, %
AstBIL (1st)	5	42.4	18	56
KarErBIL-1 (2nd)	5	40.4	18	61
Shymkent (3rd)	5	36.0	15	47
Kostanay	3	21.3	9	56
Nurmakov (4th)	5	13.8	10	40
KarErBIL-2	3	9.0	8	25
Daryn	3	6.3	9	33
Burabay	3	3.7	8	50

Team names are written as in the protocols. Places are given for the teams that reached the semifinals.

Discussion. The main result of this study is that the topic of a problem matters both for the decision to take it to the board and for the result of the defence. Combinatorics was accepted most often, geometry was accepted carefully but defended almost perfectly, and number theory was avoided and, when accepted, often not finished. In the 8th grade, number theory problems often need a key idea, such as looking at remainders or at the prime factors of both sides, and without this idea a student may have nothing to show. A geometry proof, in contrast, is either

finished or clearly not finished, so a team knows before the battle whether it can defend it. I cannot check these explanations with protocols alone, but they fit the pattern in the data.

The refusals show another side of a battle. In 61% of refusals the challenging team did not have a solution either. Under the rules, a team sometimes has to challenge on a problem that it has not solved, and the other team can use this by refusing. The data suggest that refusing was often a good decision, because in most refusals the challenging team could not defend the problem either.

For the organisers of tournaments like ours, the results give two practical ideas. First, if almost all teams refuse the same problem, as happened with the two group stage problems and the number theory problems of the final set, that problem probably did not work as part of the battle, and the jury can see it from the protocols. Second, a balanced set may need easier number theory problems than geometry problems, if the jury wants every topic to be taken to the board. The protocols also showed that the jury itself can make arithmetic mistakes: in one of 16 battles the written total was wrong by one point. Checking the sum of the rows against the total is simple and could be done after every battle.

Limitations. The data come from one league of one tournament: 8 teams, 16 battles and 30 problems. For this reason several differences, especially in acceptance, are only tendencies. The topics are connected with particular problems, so a very hard geometry problem or a very easy combinatorics problem can change the result for the whole topic. The labels of topics were made by me, and five problems could belong to two topics. Strong teams played more battles, which I tried to control with a test within teams, but with eight teams this control is limited. The protocol of the final is not available, and its score is based on the words of a member of the jury. Finally, I was a member of the jury myself, and although the points were given by the whole jury, my knowledge of the tournament may influence how I interpret the results.

Conclusion. Using the jury protocols of 16 battles of the 8th-grade league of Karagandy Mathbattles 2026, I found that teams accepted combinatorics problems most often (68%) and number theory problems least often (38%). Accepted geometry problems were almost always solved completely (mean 10.9 of 12 points), while accepted number theory problems received only 6.4 points on average. In 61% of refusals the challenging team had no solution either. Stronger teams accepted challenges more often. These results show that the protocols of a mathematical battle are a useful source of data about how students judge their own solutions, which is a part of metacognition [6]. The next step would be to repeat the analysis on the 7th-grade league of the same tournament and on other tournaments, to see whether number theory is avoided there as well.

Author contribution and use of AI. The author was one of the organisers and a member of the jury of the tournament, collected the photos of the protocols and the problem sets, transcribed all 16 battles into the table, found and corrected the errors in it, labelled the topics of the problems and made the first summary table by topics. The author's mentor checked the table against the protocols, carried out the statistical analysis, made the figures and helped write and edit the English text. Generative AI tools were used to help with editing and translation. Every number in the paper was calculated from the table by the analysis code and checked by the author's mentor.

References:

1. Kenderov P.S. Competitions and mathematics education // Proceedings of the International Congress of Mathematicians, Madrid, 2006. Zürich: European Mathematical Society, 2007. P. 1583-1598. DOI: 10.4171/022-3/76.

2. Kenderov P.S. Mathematics competitions: an integral part of the educational process // ZDM - Mathematics Education. 2022. Vol. 54, No. 5. P. 983-996. DOI: 10.1007/s11858-022-01348-4.

3. Fomin D., Genkin S., Itenberg I. Mathematical Circles (Russian Experience). Providence, RI: American Mathematical Society, 1996. (Mathematical World, Vol. 7). Chapter 18: Mathematical contests. DOI: 10.1090/mawrld/007.

4. Правила математического боя. Схема матбоя. Всероссийская смена «Юный математик», 2023. URL: https://adygmath.ru/content/files/umt23/pravila.pdf (accessed: 03.10.2026).

5. Правила математических боев. Творческая лаборатория «2×2», 2017. URL: https://mathbaby.ru/olympiads/tyumka/2017/pravila-matematicheskih-boev (accessed: 03.10.2026).

6. Schoenfeld A.H. Learning to think mathematically: problem solving, metacognition, and sense making in mathematics (reprint) // Journal of Education. 2016. Vol. 196, No. 2. P. 1-38. DOI: 10.1177/002205741619600202.

7. International Mathematical Olympiad. IMO 2023 Shortlisted Problems with Solutions. URL: https://www.imo-official.org/problems/IMO2023SL.pdf (accessed: 03.10.2026).

8. Good P. Permutation, Parametric and Bootstrap Tests of Hypotheses. 3rd ed. New York: Springer, 2005. DOI: 10.1007/b138696.

9. Kruskal W.H., Wallis W.A. Use of ranks in one-criterion variance analysis // Journal of the American Statistical Association. 1952. Vol. 47, No. 260. P. 583-621. DOI: 10.1080/01621459.1952.10483441.

УДК 373.2.016:340.1

Базарбаева Айнур Бектуровна
Магистр, к.ю.н.
Институт педагогики и психологии
Казахский Национальный Женский Педагогический Университет
(Казахстан, Алматы)

ДЕТСКИЙ ФОЛЬКЛОР И ПОДВИЖНЫЕ ИГРЫ КАК ИНСТРУМЕНТЫ ФОРМИРОВАНИЯ ОСНОВ ПРАВОВОЙ КУЛЬТУРЫ У ДЕТЕЙ СТАРШЕГО ДОШКОЛЬНОГО ВОЗРАСТА

Аннотация: В статье рассматриваются научно-методические основы формирования первичных правовых представлений и нормативного поведения у детей предшкольного возраста (5–6 лет). Автор обосновывает дефицит традиционных словесно-дидактических методов при обучении дошкольников правилам поведения и предлагает альтернативный подход, основанный на интеграции малых жанров казахского детского фольклора и методики физического воспитания (подвижных игр). В контексте исследования подвижная игра концептуализируется как первичная правовая модель, в которой игровые правила выступают аналогом социального законодательства. Анализируются оздоровительное, образовательное и воспитательное направления физической культуры в дошкольном образовании. Предложен комплексный метод использования сказкотерапии и народных игр для интериоризации детьми понятий справедливости, честности и уважения к границам личности.

Ключевые слова: дошкольное образование, правовое воспитание, детский фольклор, подвижные игры, арт-педагогика, предшкольная группа, правила игры, интериоризация норм.

Модернизация системы дошкольного образования в Республики Казахстан в соответствии с Государственным общеобязательным стандартом (ГОСО РК) требует поиска инновационных подходов к социально-эмоциональному и ценностному развитию личности ребенка. На данном возрастном этапе закладываются основы нормативного поведения, формируются первичные представления о справедливости, правах и обязанностях человека в социуме.

Традиционные подходы к правовому просвещению, заимствованные из школьной педагогики (беседы, лекции, сухие нотации), демонстрируют крайнюю неэффективность в работе с дошкольниками из-за наглядно-образного характера их мышления и психологической неготовности к восприятию абстрактных юридических категорий. Как подчеркивают современные исследователи, в системе дошкольного образования Казахстана назрела острая необходимость внедрения гибких интерактивных технологий взамен устаревших авторитарных подходов [2]. Возникает объективное научное противоречие между необходимостью формирования правовой культуры у старших дошкольников и дефицитом адекватных возрасту игровых технологий.

Настоящая статья призвана доказать, что эффективными инструментами трансляции правовых ценностей могут выступать малые жанры детского фольклора, аккумулированные в академических фондах (в частности, наследии Института литературы и искусства им. М. О. Ауэзова), в сочетании с методикой физического воспитания и подвижных игр.

Методологическую основу исследования составили культурно-историческая концепция, рассматривающая взаимосвязь игры и психического развития ребенка [1], фундаментальная теория игровой деятельности [4], а также современные западные концепции арт-педагогики и игровых технологий в раннем детстве [6].

В качестве эмпирических и теоретических методов исследования были использованы:

1. Контент-анализ и герменевтический метод: изучение академических фондов казахского детского фольклора (сказок, притч, бытовых речевок, обрядовых песен) с целью выявления в них дидактического потенциала, содержащего правовые смыслы (категории чести, справедливости, права собственности, защиты слабых) [3].
2. Моделирование педагогического процесса: проектирование структуры физкультурных занятий в ДОО с интеграцией нормативно-правового компонента в структуру подвижных игр.
3. Метод педагогического наблюдения за динамикой освоения правил детьми старшего дошкольного возраста в ходе рассредоточенной практики.

В ходе исследования было установлено, что методика физического воспитания детей дошкольного возраста обладает уникальным, латентным потенциалом для формирования правосознания. Традиционно физкультура в дошкольном образовании классифицируется по упражнениям (гимнастика, основные виды движений — ОВД, спортивные элементы) и рассматривается через призму трех ключевых направлений: оздоровительного, образовательного и воспитательного.

Именно воспитательное и образовательное направления физической культуры пересекаются с задачами правового воспитания. Центральным элементом этого пересечения выступает подвижная игра.

С научной точки зрения любая подвижная игра представляет собой первую доступную ребенку модель правового поля, что подтверждается классическими трудами по психологии игры [4]. Игра не может существовать без правил. Правила в игре выполняют функцию закона, а подчинение правилу ради сохранения игровой ситуации является высшей формой произвольного поведения дошкольника, описанной в базовых психологических исследованиях [1].

Нами выделены три последовательные стадии интериоризации правовых норм через игровую деятельность в дошкольной педагогике:

1. Стадия ознакомления. На данном этапе воспитатель озвучивает правила игры («не забегать за черту», «замереть по сигналу»). В правовом контексте это эквивалентно ознакомлению граждан с нормативно-правовым актом (Законом).
2. Стадия контроля (судейства). Дети начинают активно следить за выполнением правил сверстниками, самостоятельно выявляя нарушителей. Это формирует первичное правовое контролирующее поведение и институт честного судейства.
3. Стадия принятия (интериоризации). Происходит осознанное ограничение собственных импульсов ради принципа «Честной игры» (Fair Play), что напрямую развивает социальный интеллект ребенка [6]. На этой стадии развивается устойчивое законопослушное поведение и уважение к чужим личным границам.

Параллельно с физической активностью важнейшим элементом выступает когнитивно-речевая база — детский фольклор [3]. Изучение структуры и генезиса казахской народной литературы позволяет успешно интегрировать сказкотерапию в

структуру занятий. Герои фольклора (например, Алдар Косе, Жиренше-шешен) демонстрируют ребенку примеры разрешения споров, торжества справедливости и защиты прав личности, что закрепляется последующим проигрыванием этих сюжетов в ролях. Огромное значение здесь имеют фундаментальные исследования, раскрывающие эстетическую культуру, ценностные ориентации и духовные истоки казахского слова [5]. Таким образом, фольклор дает содержательную часть (смысл нормы), а подвижная игра — практическую форму (опыт выполнения нормы).

Проведенный теоретико-методологический анализ позволяет сделать вывод, что формирование основ правовой культуры у детей предшкольного возраста наиболее эффективно протекает в рамках интегративного подхода, объединяющего арт-педагогику (детский фольклор) и методику физического воспитания. Подвижная игра выступает естественным тренажером законопослушного поведения, где ребенок на практике осознает неотвратимость последствий за нарушение правил и ценность справедливости.

Практическая значимость исследования заключается в возможности использования предложенных классификаций и подходов для разработки методических пособий для воспитателей предшкольных организаций и студентов педагогических вузов.

Список литературы:

1. Выготский, Л. С. (2016). Игра и ее роль в психическом развитии ребенка. Книга по Требованию.
2. Канай, Г. А. (2023). Инновационные подходы в системе дошкольного образования Казахстана. Вестник Казахского национального женского педагогического университета, (2), 45-52.
3. Матыжанов, К. С. (2019). Казахский детский фольклор: генезис, структура и поэтика. Институт литературы и искусства им. М. О. Ауэзова.
4. Эльконин, Д. Б. (2021). Психология игры (2-е изд.). Владос.
5. Bazarbayev, M. (2005). Estetikalyk madeniet zhane kazirgi dauir [Aesthetic culture and the modern era] (Vols. 1–2). Zhibek Zholy.
6. Smith, J. A., & Brown, L. (2022). Art pedagogy and game technologies in early childhood education. Central Asian Educational Studies, 7(3), 112-125.

Central Asian Scientific Journal

БИОЛОГИЯ ҒЫЛЫМДАР - БИОЛОГИЧЕСКИЕ НАУКИ - BIOLOGICAL SCIENCES

UDC 639.3.05

Bazhenova Diana

Research Scientist
LLP “Fisheries Research and Production Center”
North Kazakhstan branch
(Astana, Kazakhstan)

GROWTH DYNAMICS OF STERLET (*Acipenser ruthenus L.*) REARED IN A RECIRCULATING AQUACULTURE SYSTEM

Abstract: The study evaluated the growth dynamics of sterlet (*Acipenser ruthenus L.*) reared in a recirculating aquaculture system (RAS) at the “Aqua Ferma PV” fish farm, Kazakhstan. The production cycle was followed from egg incubation to the fingerling stage, with particular attention to body weight, growth rate and survival. During the 188-day fingerling rearing period, mean body weight increased from 15.1 to 347.5 g. Total weight gain was 332.4 g, average daily gain was 1.77 g/day, and specific growth rate was 1.67%/day. Survival during this period reached 96.7%. Water quality parameters remained relatively stable during the monitored period. The obtained results demonstrate stable growth and high survival of sterlet under the applied RAS production conditions and are generally comparable with published data for juvenile sterlet reared under controlled conditions.

Keywords: recirculating aquaculture systems, sterlet, fish growth, growth dynamics;

Introduction. Aquaculture has become an increasingly important part of the fisheries sector in Kazakhstan. Growing demand for fish products, together with the need to reduce pressure on natural fish stocks, has encouraged the development of commercial fish farming and the introduction of more intensive production technologies. This trend has been particularly noticeable in recent years. Along with conventional pond and lake-based farms, Kazakhstan is developing cage and tank culture as well as recirculating aquaculture systems (RAS). According to official data, the volume of farmed fish production in the country increased from 6.8 thousand tons in 2020 to 17.1 thousand tons in 2023; RAS and tank-based farms are also represented within the structure of the country’s aquaculture enterprises [1].

Recirculating systems are of particular interest for intensive fish production because most of the water is treated and reused within the facility. Mechanical and biological filtration, aeration or oxygenation, and temperature control make it possible to maintain relatively stable water quality throughout the production cycle. Such systems are widely used in intensive sturgeon culture because they allow closer control of temperature, oxygen regime and nitrogen compounds while reducing water exchange requirements [2,7,8]. This is especially relevant for Kazakhstan, where a sharply continental climate and large seasonal temperature fluctuations can limit the duration of outdoor fish farming.

Recent studies in Kazakhstan and elsewhere show that sturgeons can be successfully reared in recirculating systems, although production results remain strongly dependent on culture conditions. Akhmetova et al., studying Siberian sturgeon (*Acipenser baerii L.*) in a recirculating system, showed that a controlled environment can support fish production

throughout the year and considered RAS-produced juveniles suitable for subsequent transfer to tanks, ponds and cages [2]. Aidos et al. [4] demonstrated that stocking density affected growth and muscle development in early-stage Siberian sturgeon larvae, while Mocanu et al. [6] showed that dietary probiotic supplementation influenced growth performance and physiological indices of Siberian sturgeon. These findings illustrate that RAS provides a controlled production environment, but the biological response of sturgeons is determined by the management of key technological factors.

Sturgeons are valuable aquaculture species due to their biological characteristics and the high commercial value of their products [5,8]. Under intensive culture, however, their performance depends on more than the type of production system itself. Water temperature and dissolved oxygen, stocking density, feeding practices, the quality of the initial stock, incubation conditions, and the transition from endogenous to exogenous feeding may all affect survival and subsequent growth [4-6,8].

Sterlet (*Acipenser ruthenus L.*) is one of the sturgeon species of interest for intensive culture. In Kazakhstan, Aitkaliyeva et al. [3] studied juvenile sterlet and Siberian sturgeon originating from the Irtysh River under recirculating conditions, with particular attention to water quality, morphological and production characteristics, stocking density, survival and growth. Nicula et al. [7] evaluated the growth performance of juvenile sterlet reared in RAS, whereas Ginayatov et al. [9] investigated the effects of oxygen concentration and stocking density on sterlet growth and development. Other studies have focused on more specific production factors: Thomas et al. [10] examined sterlet development and behaviour under enrichment and polyculture conditions, while Bavi et al. [11] assessed the effects of dietary taurine on growth, feed utilization and physiological responses of juvenile sterlet. Thus, previous research has addressed a wide range of individual factors affecting sterlet performance under controlled rearing conditions.

Despite the available studies, most published work focuses on individual technological or nutritional factors and relatively short experimental periods. Production-scale data that follow the same sterlet batch through successive rearing stages are less frequently reported. For commercial RAS, such information is important because changes in relative and absolute growth during the production cycle influence feeding, grading, stocking and biomass management. Following one production batch from incubation through the fingerling stage therefore provides a practical view of how growth and survival develop under routine farm conditions.

Therefore, the aim of the present study was to assess the growth dynamics of sterlet (*Acipenser ruthenus* L.) at different stages of rearing in a RAS, from egg incubation to the fingerling stage, and to compare the obtained growth performance with published data for sterlet reared under controlled conditions.

Materials and Methods.

RAS production system at "Aqua Ferma PV".

The study was carried out at the "Aqua Ferma PV" fish farm on a production batch of sterlet (Acipenser ruthenus L.) reared in a recirculating aquaculture system (RAS). The system included mechanical and biological filtration, UV treatment and oxygenation. Separate facilities were used for egg incubation, larval rearing and further rearing of sterlet fingerlings. The fingerling unit consisted of six circular tanks, each 3.5 m in diameter and 1.2 m in height, with a working volume of 6.5 m^3. During the fingerling rearing period, water temperature was maintained at approximately 19°C. Water quality was controlled as part of routine RAS

management: pH, nitrite and nitrate were generally measured once every three days using commercial drop test kits, while dissolved oxygen was measured daily using a Merke oximeter. The monitored parameters were compared with recommended ranges for sturgeon culture in RAS [7,8].

Main production processes of sterlet rearing.

The studied batch originated from fertilized eggs delivered to the farm in December 2025. The production cycle was followed from egg incubation to the fingerling stage, until August 2026. Farm production records were used to obtain data on egg incubation and hatching, larval survival and mortality, feeding, grading and transfer of fish, as well as control weighings performed during the rearing period.

Fertilized sterlet eggs were obtained from the “Karaganda Osetr” fish farm. The eggs were collected on 30 November 2025 and delivered to “Aqua Ferma PV” on 2 December 2025. The batch consisted of 1.5 kg of eggs, corresponding to about 100,000 eggs.

Incubation started on 2 December 2025. Four Weiss jars were used, with approximately 350 g of eggs placed in each jar. According to technology water temperature was maintained at 14°C during incubation. Methylene blue was used as a preventive treatment against fungal infection.

Egg development was monitored during the incubation period, and non-developing eggs were removed. Hatching started five days after the beginning of incubation and continued for approximately two days. After hatching, the larvae were transferred to the larval rearing unit.

Sterlet larvae were kept in separate larval rearing troughs. During the first days after hatching, the larvae remained on endogenous nutrition. Active feeding started on days 7-8 after hatching.

At the beginning of exogenous feeding, newly hatched Artemia nauplii and a commercial starter feed (Aller Aqua Silver, 0.2 mm) were used. During the first three days of active feeding, the larvae were fed up to 12 times per day. Artemia nauplii were supplied according to feed consumption. Live feed was used together with the starter diet during the initial feeding period, followed by gradual transition to the commercial diet.

The condition of the larvae, feed consumption and mortality were monitored throughout this period. The first grading was performed on 10 January 2026. Larval rearing continued until approximately 60 days of age, when the fish were transferred to the fingerling rearing unit.

At 60 days of age, sterlet fingerlings were transferred to the circular tanks of the RAS fingerling unit. The initial stocking rate was 2,000 fish per tank.

Fish growth was monitored by regular control weightings. The interval between successive weightings ranged mainly from 10 to 20 days according to the farm production schedule. Body weight recorded at each control point was used to follow the growth of the studied batch throughout the fingerling rearing period.

Commercial feeds were used during this stage, with pellet size adjusted as the fish increased in size. The pellet size increased to 2 mm in March 2026 and to 3 mm in June 2026. Grading and redistribution of fish between tanks were carried out when required during the production cycle.

Control observations continued until August 2026. Successive body weight measurements were subsequently used to assess the growth dynamics of the studied sterlet batch. The main production stages and rearing conditions of the studied sterlet batch are summarized in Table 1.

Table 1 - Main production stages and rearing conditions of sterlet at the "Aqua Ferma PV" fish farm.

Production stage	Period	Rearing facility	Feeding	Monitoring
Egg incubation	Dec 2025	4 Weiss jars; ~350 g eggs/jar	-	Egg development, hatching
Larval rearing	Dec 2025-Feb 2026	Larval rearing troughs	*Artemia* nauplii + Aller Aqua Silver, 0.2 mm	Survival, feeding, grading, body weight
Fingerling rearing	Feb-Aug 2026	6 circular tanks, 6.5 m³ each	Commercial formulated diets; pellet size adjusted to fish size	Control weighing, growth

Monitoring of fish-biological indicators.

Changes in mean body weight were used as the main indicator of growth. For each available control weighing, the date, fish age, number of individuals and mean body weight were recorded. These measurements were used to construct the growth curve from the early stages of development to the juvenile stage. During grading, the minimum and maximum body weights were recorded for each size group. For further analysis, mean body weight was calculated as the arithmetic mean of the minimum and maximum recorded values.

Based on these data, growth performance was assessed using absolute weight gain (WG), average daily gain (ADG), and specific growth rate (SGR), following approaches commonly applied in studies of sturgeon growth [4-6].

Absolute weight gain (WG) was used to determine the increase in mean body weight over a given rearing period and was calculated as:

$$WG = W_2 - W_1 \quad (1),$$

where W_1 - body weight at the beginning of the period (g), and W_2 - body weight at the end of the period (g).

Average daily gain (ADG) was calculated to estimate the average increase in body weight per day during each interval between control weightings:

$$ADG = \frac{(W_2 - W_1)}{t} \quad (2),$$

where W_1 - body weight at the beginning of the period (g), and W_2 - body weight at the end of the period (g); t - the period between measurements (days). ADG was expressed in g day^{-1}.

Specific growth rate (SGR) was used to characterize growth relative to the body weight of the fish. SGR was calculated as:

$$SGR = \frac{\ln W_2 - \ln W_1}{t} \times 100 \quad (3),$$

where W_1 - body weight at the beginning of the period (g), and W_2 - body weight at the end of the period (g); t - the period between measurements (days); ln - the natural logarithm. SGR was expressed as % day^{-1}.

Survival rate (SR) was determined separately for the main production stages as:

$$SR = \frac{N_2}{N_1} \times 100 \quad (4),$$

where N_1 - the number of fish at the beginning of the period; N_2 - the number remaining at the end of the period. Survival was expressed as a percentage (%).

Calculated growth parameters were compared with published data for sterlet and other sturgeon species reared in RAS.

Results. During the incubation period, the first sterlet larvae appeared five days after the beginning of incubation, while hatching continued for approximately two days. According to the production records, about 20,000 larvae were recorded at the swarming stage. During the transition to exogenous feeding and adaptation to the commercial diet, larval mortality was approximately 10-15%. At the first grading, performed on 10 January 2026, the mean body weight of sterlet was 0.5 g.

Further rearing was characterized by a continuous increase in body weight. At the beginning of the fingerling rearing period, approximately 12,000 fish were distributed among six rearing tanks. At the first control weighing on 10 February, body weight ranged from 14.7 to 15.5 g, with a calculated mean value of 15.1 g. By 22 March, mean body weight increased to 61.0 g, and on 20 April it reached 87.0 g. A further increase was recorded in May and June, reaching 122.5 g on 26 May and 167.5 g on 15 June.

The highest increase in body weight was observed during the summer period. Mean weight increased from 167.5 g on 15 June to 232.5 g on 8 July and reached 285.0 g by 27 July. At the last control weighing, performed on 17 August 2026, body weight ranged from 340 to 355 g, with a calculated mean value of 347.5 g (Figure 1). Thus, during the 188-day fingerling rearing period, mean body weight increased from 15.1 to 347.5 g.

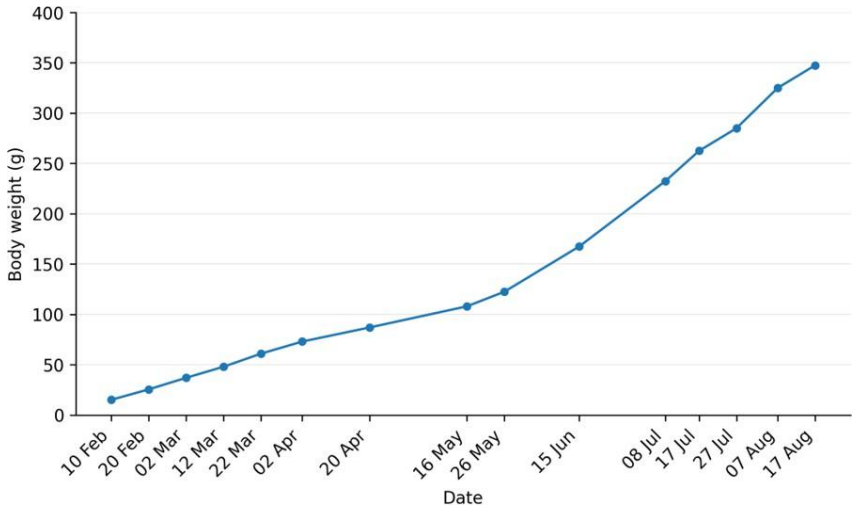


Figure 1 - Dynamics of mean body weight of sterlet (*Acipenser ruthenus L.*) during the fingerling rearing period.

Based on the recorded body weight data, weight gain (WG), average daily gain (ADG) and specific growth rate (SGR) were calculated for the fingerling rearing period. The total individual weight gain over the 188-day period was 332.4 g, corresponding to an average daily gain of 1.77 g/day. The overall SGR was 1.67%/day.

The highest SGR was recorded during the initial growth interval, reaching 5.24%/day. As body weight increased, SGR gradually declined, whereas absolute daily weight gain generally increased. The highest ADG was observed between 27 July and 7 August, reaching 3.64 g/day.

At the end of the observation period, 11,600 of the initial 12,000 fingerlings remained in the rearing unit, corresponding to a survival rate of 96.7% (Table 2).

Table 2. Growth performance of sterlet during the fingerling rearing period

Rearing period	Days	WG (g fish^{-1})	ADG (g day^{-1})	SGR (% day^{-1})
10-20 Feb	10	10.4	1.04	5.24

20 Feb-2 Mar	10	11.5	1.15	3.72
2-12 Mar	10	11.0	1.10	2.60
12-22 Mar	10	13.0	1.30	2.40
22 Mar-2 Apr	11	12.0	1.09	1.63
2-20 Apr	18	14.0	0.78	0.97
20 Apr-16 May	26	21.0	0.81	0.83
16-26 May	10	14.5	1.45	1.26
26 May-15 Jun	20	45.0	2.25	1.56
15 Jun-8 Jul	23	65.0	2.83	1.43
8-17 Jul	9	30.0	3.33	1.35
17-27 Jul	10	22.5	2.25	0.82
27 Jul-7 Aug	11	40.0	3.64	1.19
7-17 Aug	10	22.5	2.25	0.67
Overall	**188**	**332.4**	**1.77**	**1.67**

Water quality parameters remained relatively stable during the monitored period. Dissolved oxygen ranged from 6.5 to 7.0 mg/L, while pH varied between 6.0 and 7.5. Nitrite concentration remained low and ranged from 0 to 0.1 mg/L. Nitrate concentration varied from 5 to 25 mg/L, with a gradual decrease towards the end of the monitored period. Water temperature during the fingerling rearing period was maintained at approximately 19°C.

Discussion. The growth performance obtained in the present study was generally comparable with data reported for sterlet under controlled rearing conditions. Ginayatov et al. [9] reported an SGR of 1.47%/day and an average daily gain of 1.43 g/day in sterlet reared in RAS at a stocking density of 40 kg/m³. In the present study, the corresponding values for the entire fingerling rearing period were 1.67%/day and 1.77 g/day, respectively.

Similar final body weights were reported by Thomas et al. [10] in a 90-day experiment with sterlet. Fish with an initial body weight of 132.6-133.1 g reached 324.7-337.1 g at the end of the experiment, with SGR values ranging from 0.99 to 1.04%/day. In the present study, sterlet reached a mean body weight of 347.5 g by the final control weighing. Although the duration of rearing and initial body weight differed between the studies, the final body weights were similar.

Growth rate varied considerably during the rearing period. The highest SGR in the present study was recorded during the first interval (5.24%/day), followed by a gradual decrease as fish body weight increased. In contrast, absolute daily weight gain was higher during the later stages of rearing, reaching a maximum of 3.64 g/day between 27 July and 7 August. Differences in growth rate during sterlet culture may be related to fish size, duration of the rearing period, feeding conditions and environmental parameters. Bavi et al. [11], for example, reported SGR values of 1.22-1.55%/day in juvenile sterlet with an initial body weight of approximately 14 g during a 70-day feeding trial.

Survival during the fingerling rearing period was 96.7%. This value was also comparable with results reported for juvenile sterlet under controlled conditions. Bavi et al. [11] recorded survival rates ranging from 88.89 to 97.23%, depending on the experimental diet. The high survival recorded in the present study, together with the observed increase in body weight, indicates stable performance of sterlet during the fingerling rearing period.

Conclusion. The study showed that the applied RAS rearing conditions supported stable growth and high survival of sterlet during the fingerling stage. A consistent growth pattern was observed: as body weight increased, relative growth rate declined, whereas absolute daily gain generally increased. The obtained results support the use of RAS for intensive rearing of sterlet

juveniles and provide production-scale information that can be used for feeding, grading and biomass management.

References

1. Fisheries Committee of the Ministry of Ecology and Natural Resources of the Republic of Kazakhstan [Электронный ресурс]. - URL: https://www.gov.kz/memleket/entities/fish (дата обращения: 26.08.2026).
2. Akhmetova B., Nurzhanova K., Satieva K., Kurmangali L. Research of breeding and development of sturgeon fishes in recirculating aquaculture systems // Science and Education. - 2023. - Vol. 3, No. 2 (71). - P. 52-59. - DOI: 10.52578/2305-9397-2023-2-3-52-59.
3. Aitkaliyeva A. A., Asylbekova S. Zh., Shukurov M. Zh., Sitakhmetova G. K., Bulavina N. B., Aubakirova M. O., Ablaysanova G. M. Peculiarities of growing junior sturgeon fish from the Irtysh River in recirculating aquaculture systems // Science and Education. - 2025. - Vol. 5, No. 3 (80). - P. 233-244. - DOI: 10.52578/2305-9397-2025-3-5-233-244.
4. Aidos L., Cafiso A., Serra V., Vasconi M., Bertotto D., Bazzocchi C., Radaelli G., Di Giancamillo A. How different stocking densities affect growth and stress status of *Acipenser baerii* early-stage larvae // Animals. - 2020. - Vol. 10, No. 8. - Art. 1289. - DOI: 10.3390/ani10081289.
5. Wiszniewski G., Jarmołowicz S., Hassaan M. S., Soaudy M. R., Kamaszewski M., Szudrowicz H., Terech-Majewska E., Pajdak-Czaus J., Wiechetek W., Siwicki A. K. Beneficial effects of dietary papain supplementation in juvenile sterlet (*Acipenser ruthenus*): growth, intestinal topography, digestive enzymes, antioxidant response, immune response, and response to a challenge test // Aquaculture Reports. - 2022. - Vol. 22. - Art. 100923. - DOI: 10.1016/j.aqrep.2021.100923.
6. Mocanu E. E., Savin V., Popa M. D., Dima F. M. The effect of probiotics on growth performance, haematological and biochemical profiles in Siberian sturgeon (*Acipenser baerii* Brandt, 1869) // Fishes. - 2022. - Vol. 7, No. 5. - Art. 239. - DOI: 10.3390/fishes7050239.
7. Nicula M., Szelei Z.-T., Păcală N., Bura M., Simiz E., Erina S. Assessment of growth performances of juvenile sterlet (*Acipenser ruthenus*) raised in a recirculating aquaculture system (RAS) // Scientific Papers: Animal Science and Biotechnologies. - 2015. - Vol. 48, No. 2. - P. 108-114.
8. Chebanov M. S., Galich E. V. Sturgeon hatchery manual. - Ankara : FAO, 2011. - 303 p. - (FAO Fisheries and Aquaculture Technical Paper ; No. 558).
9. Ginayatov N., Shukurov M., Shumeyko D., Sariyev B., Arystangalieva V., Abdessan R., Brigida A., Albekov A., Bexultan A. Effect of different oxygen concentrations and stocking density on the growth and development of *Acipenser ruthenus* in a recirculating aquaculture system // Asian Journal of Agriculture and Biology. - 2026. - Art. e2025257. - DOI: 10.35495/ajab.2025.257.
10. Thomas M., Reynaud J.-G., Ledoré Y., Pasquet A., Lecocq T. Enrichment in a fish polyculture: does it affect fish behaviour and development of only one species or both? // Applied Sciences. - 2022. - Vol. 12, No. 7. - Art. 3674. - DOI: 10.3390/app12073674.
11. Bavi Z., Zakeri M., Mousavi S. M., Yavari V. Effects of dietary taurine on growth, body composition, blood parameters, and enzyme activities of juvenile sterlet (*Acipenser ruthenus*) // Aquaculture Nutrition. - 2022. - Vol. 2022. - Art. 1713687. - 13 p. - DOI: 10.1155/2022/1713687.

ТУРИСТИЧЕСКИЕ И СПОРТИВНЫЕ НАУКИ - ТУРИСТІК ЖӘНЕ СПОРТ ҒЫЛЫМДАРЫ - TOURISM AND SPORTS SCIENCES

ӘОЖ 338.48

Төлеп Жазира Маратқызы
магистрант
Ғылыми жетекші: Тагманов Урманбек Сарманович
PhD., аға оқытушы
Халықаралық туризм және меймандостық университеті
(Түркістан қ., Қазақстан)

ГАСТРОНОМИЯЛЫҚ ФЕСТИВАЛЬДЕРДІҢ ТУРИСТІК ДЕСТИНАЦИЯ ТАРТЫМДЫЛЫҒЫН ҚАЛЫПТАСТЫРУДАҒЫ РӨЛІ МЕН ДАМУ ҮРДІСТЕРІ

Аңдатпа: Мақалада гастрономиялық фестивальдердің туристік дестинация тартымдылығын қалыптастырудағы рөлі Түркістан қаласы мысалында бағаланады. Зерттеудің мақсаты гастрономиялық фестивальдің дестинация имиджіне әсер ету тетіктерін айқындау, Түркістандағы фестивальдік гастрономияның 2022-2025 жылдардағы даму бағытын сипаттау және оның туристік нәтижесін бағалауға жарамды көрсеткіштерді ұсыну болды. Эмпирикалық база ретінде ресми мемлекеттік ресурстардағы Gastro Fest Turkistan іс-шаралары, Түркістанның туристік оқиғалар күнтізбесі және өңірлік туризм статистикасы пайдаланылды. Зерттеуде салыстырмалы кейстік талдау, контент-талдау және сипаттамалық статистика әдістері қолданылды. Ресми хабарламаларда Gastro Fest Turkistan фестиваліне 2022 жылы 2 мыңнан астам, 2023 жылы 10 мыңнан астам келуші қатысқаны, ал қоғамдық тамақтану нысандарының саны 25-тен астам мекемеден 30-ға жуық мекемеге дейін кеңейгені көрсетілген. Сонымен бірге гастрокартаның енгізілуі және кейінгі жылдары гастрономиялық мазмұнның этномәдени және ойын-сауық іс-шараларымен бірігуі байқалады. Ашық деректер нақты туристік әсерді себептік тұрғыдан өлшеуге жеткіліксіз болғандықтан, мақалада фестиваль тиімділігін келушінің географиясы, түнеу ұзақтығы, шығыны, қанағаттануы, қайта келу ниеті және жергілікті тағамды тану деңгейі арқылы бағалау ұсынылады.

Түйін сөздер: гастрономиялық туризм, гастрономиялық фестиваль, туристік дестинация, Түркістан, жергілікті тағам, дестинация имиджі, оқиғалық туризм.

Қазіргі туризмде дестинацияның тартымдылығы тарихи нысандардың санымен ғана емес, турист алатын тұтас тәжірибенің сапасымен айқындалады. Осы тұрғыдан гастрономия жергілікті мәдениетті күнделікті тұтыну тәжірибесімен байланыстыратын маңызды ресурс болып саналады. M. P. Leal Londoño, A. Georgescu-Paquin және J. Arcos-Pumarola тағам фестивальдерінің дестинация бейнесін қалыптастыруға қатысатынын және ұйымдастырушы ұсынған образ бен турист қабылдайтын образдың сәйкестігі шешуші мәнге ие екенін көрсетеді [1].

Гастрономиялық фестиваль жергілікті тағамды, қоғамдық тамақтану кәсіпорындарын, мәдени бағдарламаны және туристің оқиғаға қатысу тәжірибесін бір кеңістікте біріктіреді. A. T. T. Tran және бірлескен авторлардың шолуында фестиваль саласы, қызмет көрсету, келуші қанағаттануы, гастрономиялық тәжірибе және адалдық негізгі зерттеу бағыттары ретінде белгіленген [2]. Сондықтан фестивальді бір реттік көпшілік шара емес, дестинацияның туристік өнімі ретінде бағалау орынды.

Тарихи-мәдени дестинацияларда жергілікті асхана аумақтың бейнесін толықтыратын мазмұндық ресурсқа айналады. N. Huete-Alcocer мен R. D. Hernandez-Rojas дәстүрлі асханамен қанағаттану туристің дестинацияны когнитивтік және эмоциялық қабылдауына, сондай-ақ қайта келу және ұсыну ниетіне ықпал ететінін анықтаған [3]. Түркістан жағдайында гастрономиялық оқиға тарихи және зияраттық брендті алмастырмай, оны жергілікті тағам мен қонақжайлық тәжірибесі арқылы тереңдетуі тиіс.

Фестивальдің құндылығы мәзір ассортиментімен шектелмейді. F. C. Hsu, E. Agyeiwaah және L. I. L. Chen гастрономиялық фестивальде тәжірибелік құндылық, шынайылықты қабылдау, эмоциялық реакция, қанағаттану және адалдық өзара байланысты екенін көрсетеді [4]. Демек, тағамның шығу тегі, дайындау тәсілі және жергілікті мәдениетпен байланысы туристік тәжірибенің бір бөлігі ретінде ұсынылуы қажет.

Келушінің іс-шараға қатысу деңгейі де нәтижеге әсер етеді. H. Choo, D. B. Park және J. F. Petrick жергілікті тағам фестивальдерінде сапа мен қанағаттанудың адалдықпен байланысын, ал белсенді қатысудың бұл байланысты күшейтетінін көрсеткен [5]. Сондықтан дегустация, шеберлік сабағы, аспазбен өзара әрекет және тағам тарихын түсіндіру сияқты форматтар маңызды.

Соңғы зерттеулер гастрономиялық фестивальдің дестинация имиджімен байланысын нақтылай түседі. 2026 жылы Сербиядағы Belmužijada фестивалі бойынша жүргізілген зерттеуде фестиваль сапасы, қанағаттану, дестинация имиджі және қайта келу ниеті арасында статистикалық мәнді байланыстар анықталған [6]. Бұл нәтижелер Түркістан үшін фестиваль тиімділігін келушілер санымен ғана емес, туристік мінез-құлық көрсеткіштерімен бағалау қажеттігін негіздейді.

Зерттеудің мақсаты Түркістандағы гастрономиялық фестивальдердің туристік дестинация тартымдылығын қалыптастырудағы рөлін бағалау және 2022-2025 жылдардағы даму үрдістерін жүйелеу болып табылады. Зерттеу нысаны - Түркістан қаласында өткізілген гастрономиялық мазмұны бар оқиғалар, ал зерттеу пәні - фестиваль форматының дестинация имиджімен, туристік тәжірибемен және жергілікті туристік өніммен байланысы.

Зерттеу салыстырмалы кейстік талдауға негізделді. Теориялық бөлікте гастрономиялық фестиваль, жергілікті асхана, дестинация имиджі және туристік адалдық жөніндегі рецензияланған еңбектер жүйеленді. Эмпирикалық бөлікте 2022 және 2023 жылдардағы Gastro Fest Turkistan туралы ресми хабарламалар, 2024 жылғы туристік іс-шаралар бағдарламасы, 2025 жылғы гастрономиялық оқиғалар және өңірлік туризм статистикасы мазмұндық және салыстырмалы талдау арқылы қарастырылды.

Ресми деректерді интерпретациялауда олардың шектеулері ескерілді. 2022 жылғы аудитория «2 мыңнан астам», ал 2023 жылғы аудитория «10 мыңнан астам» деп жарияланған. Бұл сандар нақты есеп емес, төменгі шектер болғандықтан, олар негізінде дәл өсім коэффициентін есептеу дұрыс емес. Сонымен қатар ашық дереккөздер келушілердің турист немесе жергілікті тұрғын екенін және түнеу фактісін бөлмейді. Сондықтан мақалада фестивальге қатысушылар саны туристік ағыммен теңестірілмейді.

Қазақстан Республикасының туристік саласын дамытудың 2023-2029 жылдарға арналған тұжырымдамасында туристік өнімді әртараптандыру, дестинациялардың бәсекеге қабілеттілігін арттыру және сапалы қызметті дамыту басымдық ретінде белгіленген [7]. Түркістандағы гастрономиялық фестивальдер осы міндеттерге сәйкес

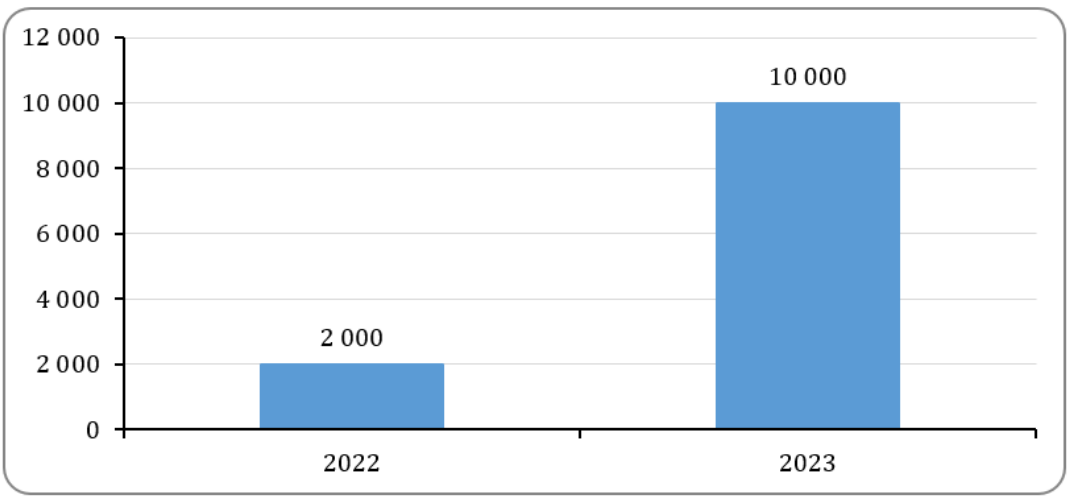
тарихи-зияраттық туризмді тәжірибелік мазмұнмен толықтырып, шағын және орта бизнесті туристік құндылық тізбегіне тартуға мүмкіндік береді.

Түркістандағы дербес гастрономиялық фестиваль форматы 2022 жылы Gastro Fest Turkistan іс-шарасымен айқындалды. Ресми мәліметте фестивальге 2 мыңнан астам адам қатысып, 25-тен астам мейрамхана мен дәмхана ұлттық тағамдар, жергілікті деликатестер, көше тағамдары, десерттер мен сусындар ұсынғаны көрсетілген. Бағдарламаға концерт, ойындар, байқаулар және шеберлік сабақтары енгізіліп, Түркістанның алғашқы гастрокартасы таныстырылған [8]. Гастрокарта фестиваль тәжірибесін қаланың тұрақты гастрономиялық инфрақұрылымымен байланыстыратын маңызды элемент болды.

2023 жылы Gastro Fest Turkistan Наурыз мерекесі аясында жалғасты. Ресми жарияланым бойынша 30-ға жуық мейрамхана мен дәмхана қатысып, 10 мыңнан астам қонақ келген. Бағдарлама ұлттық тағамдар, жергілікті деликатестер, аспаздардың шеберлік сабақтары және ойын-сауық мазмұнымен толықтырылған. Open Turkistan жаңартылған гастрокартаны және жеңілдік купондарын ұсынған [9]. Бұл тәжірибе фестивальдің жергілікті қоғамдық тамақтану желісімен байланысының күшейгенін көрсетеді.

1-кесте - Gastro Fest Turkistan фестивалінің 2022 және 2023 жылдардағы негізгі сипаттамалары

Көрсеткіш	2022 жыл	2023 жыл	Талдаулық мәні
Келушілер саны	2 мыңнан астам	10 мыңнан астам	Ресми жарияланған төменгі шектер; нақты өсім коэффициентін есептеуге жеткіліксіз
Қатысушы қоғамдық тамақтану нысандары	25-тен астам	30-ға жуық	Фестивальге тартылған бизнес шеңберінің кеңеюі байқалады
Негізгі гастрономиялық мазмұн	Ұлттық және жергілікті тағамдар, көше тағамдары	Ұлттық тағамдар, жергілікті деликатестер, көше тағамдары, сусындар	Жергілікті және көпшілікке таныс форматтарды үйлестіру



1-сурет - Gastro Fest Turkistan фестивалі бойынша ресми жарияланған келушілер санының төменгі шектері, адам

Диаграммада ресми хабарламаларда көрсетілген төменгі шектер ғана берілген. 2 000 және 10 000 мәндерінің айырмасы аудитория ауқымының кеңейгенін көрсетеді, бірақ нақты бес еселік өсім туралы қорытынды жасауға мүмкіндік бермейді. Ғылыми тұрғыдан маңызды белгі - фестивальдің қайталануы, қатысушы кәсіпорындар санының артуы және гастрокарта арқылы оқиғаның қалалық гастрономиялық ұсыныспен байланысының күшеюі.

2024 жылы Түркістанның Түркі мемлекеттерінің туристік астанасы мәртебесі оқиғалық туризмді кеңейтуге мүмкіндік берді. Ресми жоспар бойынша қалада 100-ге жуық халықаралық, республикалық және облыстық іс-шара өткізу көзделіп, олардың қатарына «Түркі әлемінің тағамдары - Тәтті» гастрофестивалі мен «Көне түркі дәстүрі» этнофестивалі енгізілген [10]. Бұл гастрономиялық мазмұнның жеке фестиваль шеңберінен шығып, дестинацияның жалпы оқиғалық портфеліне енгенін көрсетеді.

2025 жылғы тәжірибеде гастрономиялық мазмұн гибридтік форматтарға біріктірілді. Түркістандағы су және жарық шоуымен байланысты іс-шара аясында ұйымдастырылған Taiqazan гастрофестивалінде ұлттық тағамдар, сусындар және аспаздардың шеберлік сабақтары ұсынылған [11]. Мұндай формат аудиторияны кеңейтеді, алайда гастрономиялық компонент жалпы бағдарламада көмескіленбеуі үшін жергілікті тағам мен Түркістанға тән мәдени символиканың айқын болуы қажет.

Өңірлік статистика гастрономиялық өнімді дамытуға жеткілікті сұраныс базасы бар екенін көрсетеді. 2025 жылдың қорытындысы бойынша Түркістан облысының орналастыру орындарында 347,6 мың турист демалған, оның 22 563-і шетелдік турист болған; бұл көрсеткіш алдыңғы жылмен салыстырғанда 14,7 пайызға жоғары. Тарихи-мәдени нысандарға келушілер саны 1 млн адамға жеткен [12]. Бұл деректер фестивальдің себептік әсерін дәлелдемейді, бірақ гастрономиялық ұсынысты қалыптасқан туристік ағыммен байланыстыруға мүмкіндік бар екенін көрсетеді.

Салыстырмалы талдау төрт үрдісті айқындады. Біріншісі - фестивальдің қайталануы және аудиториясының кеңеюі. Екіншісі - гастрокарта арқылы қоғамдық тамақтану желісімен байланысының күшеюі. Үшіншісі - гастрономиялық мазмұнның этномәдени және ойын-сауық форматтарына енуі. Төртіншісі - гастрономияның қаланың оқиғалық коммуникациясының тұрақты элементіне айналуы. Бұл үрдістер фестивальдік гастрономияның жеке оқиғадан жүйелі дестинациялық өнімге ауысу әлеуетін көрсетеді.

Алайда фестиваль танымалдығы мен дестинация тартымдылығын тең ұғым ретінде қарастыруға болмайды. Көпшілік іс-шараға жергілікті тұрғындардың қатысуы оның әлеуметтік маңызын арттырғанымен, туристік нәтиже туралы жеке қорытынды бермейді. Дестинациялық әсерді бағалау үшін өңірден тыс келушілер үлесі, түнеу саны, орташа шығын, сапардан кейінгі қанағаттану, қайта келу және ұсыну ниеті сияқты көрсеткіштер қажет.

Түркістан үшін мәзірдің аумақтық сәйкестігі стратегиялық мәнге ие. Кез келген ұлттық тағам автоматты түрде қаланың гастрономиялық брендіне айналмайды. Тағамның өңірмен тарихи, мәдени немесе өндірістік байланысы түсіндіріліп, атауы, шығу тегі және дайындау тәсілі жөнінде қысқа ақпарат берілуі тиіс. Сонда гастрономиялық ұсыныс жай тұтыну нысанынан Түркістанды есте сақтауға көмектесетін мәдени тәжірибеге айналады.

Фестивальдің экономикалық әсері жергілікті кәсіпорындардың туристік шығыннан үлес алуымен байланысты. Бір күндік сауда көлемі тұрақты дестинациялық нәтиженің толық өлшемі емес. Нәтиже фестивальден кейін де мейрамхана, базар, гастрокарта, жергілікті өндіруші және экскурсиялық маршрут арасындағы сұраныс сақталған жағдайда күшейеді. Сондықтан гастрокарта ақпараттық буклет қана емес, сапар бағытын ұзартып, туристік тұтынуды қала ішінде бөлетін құрал ретінде қарастырылуы тиіс.

Түркістанның тарихи және зияраттық сипаты гастрономиялық фестивальдің мазмұнына қосымша талап қояды. Гастрономиялық өнім кесене мен тарихи орталыққа бәсекелес емес, негізгі мәдени мотивті толықтыратын тәжірибе болуы қажет. Тағам дайындау технологиясы, дастархан әдебі, маусымдық дәстүр және жергілікті қонақжайлық туралы түсіндірме туристің мәдени тәжірибесін тереңдетіп, дестинацияның біртұтас бейнесін сақтайды.

Фестиваль тиімділігін бағалау тек келушілер санымен шектелмеуі тиіс. Ұйымдастырушылар жыл сайын бірдей көрсеткіштер жүйесін қолданса, нәтижені динамикада салыстыруға болады. Базалық индикаторларға аудиторияның шығу географиясы, сапар мақсаты, түнеу саны, орташа шығын, қанағаттану, жергілікті тағамды қабылдау, қайта келу және ұсыну ниеті жатады. Деректерді қысқа сауалнама, QR-код, кәсіпорындардың агрегатталған сату көрсеткіштері және орналастыру орындарының статистикасы арқылы жинауға болады.

2-кесте - Гастрономиялық фестивальдің дестинациялық әсерін бағалау көрсеткіштері

Бағалау бағыты	Негізгі көрсеткіш	Түсіндіру
Аудитория құрылымы	Өңірден тыс және шетелдік келушілер үлесі	Фестивальдің жергілікті мерекеден туристік тарту құралына айналуын көрсетеді
Сапар ұзақтығы	Орташа түнеу саны	Оқиғаның қалада қалу уақытын ұзарта алуын сипаттайды
Экономикалық әсер	Бір келушінің орташа шығыны, қатысушы бизнестің қосымша айналымы	Туристік сұраныстың жергілікті экономикаға берілуін көрсетеді
Тәжірибе сапасы	Қанағаттану, сервис пен бағдарламаны бағалау	Фестивальдің тікелей сапасын өлшейді
Дестинация имиджі	Түркістанды жергілікті тағаммен байланыстыру деңгейі	Гастрономиялық мазмұнның қала бейнесіне енуін көрсетеді
Мінез-құлық ниеті	Қайта келу және өзгелерге ұсыну ниеті	Фестиваль әсерінің сапардан кейін сақталуын сипаттайды
Жергілікті кірігу	Жергілікті тағам, өндіруші және шикізат үлесі	Гастрономиялық туризмнің аумақтық ерекшелікке негізделуін көрсетеді

Ұсынылған көрсеткіштер қоғамдық танымалдықты нақты туристік нәтижеден ажыратуға мүмкіндік береді. Егер келушілер саны өсіп, бірақ өңірден тыс аудитория үлесі, түнеу немесе жергілікті бизнестегі шығын өзгермесе, әсер негізінен оқиғалық деңгейде қалуы ықтимал. Ал осы көрсеткіштер дестинация имиджі және қайта келу ниетімен қатар жақсарса, фестивальдің дестинациялық функциясы күшейді деген қорытынды жасауға болады.

Зерттеу нәтижелері Түркістандағы гастрономиялық фестивальдерді дамытудың үш бағытын негіздейді. 1. Gastro Fest Turkistan атауын жыл сайынғы негізгі гастрономиялық оқиға ретінде тұрақтандырып, өткізу мерзімі мен мазмұндық өзегін алдын ала жариялау қажет. Тұрақты күнтізбе фестивальді сапарды жоспарлауға ықпал ететін дестинациялық бренд активіне айналдырады.

2. Фестиваль мазмұны Түркістанға тән жергілікті тағам мен оның тарихына негізделуі тиіс. Қатысушы кәсіпорындар үшін жергілікті өнім үлесі, тағамның шығу тегі, баға ашықтығы, санитариялық талап және қазақ, орыс, ағылшын тілдеріндегі қысқа сипаттама бойынша ортақ стандарт енгізу орынды. Фестивальден кейін туристі гастрокартадағы мейрамханаларға, базарларға және жергілікті өнім сататын нүктелерге бағыттау бір күндік тәжірибені тұрақты маршрутқа айналдырады.

3. Фестиваль нәтижесін бағалаудың бірыңғай хаттамасын енгізу қажет. Келушілердің саны, тұрғылықты жері, сапар мақсаты, түнеу ұзақтығы, орташа шығын, қанағаттану, қайта келу және ұсыну ниеті жыл сайын бір әдістememен өлшенгенде ғана оқиғаның туристік және экономикалық нәтижесін салыстыруға, бағдарламаның әлсіз бөліктерін анықтауға және келесі фестивальді дерекке сүйеніп жоспарлауға болады.

Зерттеудің негізгі шектеуі ашық ресми деректердің агрегатталған сипатымен байланысты. Қолда бар материалдар қатысушының турист немесе жергілікті тұрғын екенін, қай өңірден келгенін, қалада түнеуін және қанша қаражат жұмсағанын анықтауға мүмкіндік бермейді. Сондықтан мақалада Gastro Fest Turkistan фестивалінің туристік ағымға себептік әсері туралы тұжырым жасалмайды. Келесі эмпирикалық кезеңде қатысқан және қатыспаған туристерді салыстыру, сондай-ақ дестинация имиджі, гастрономиялық шынайылық, қанағаттану және қайта келу ниеті арасындағы байланысты регрессиялық немесе құрылымдық модель арқылы тексеру орынды.

Жүргізілген талдау гастрономиялық фестиваль Түркістанның туристік тартымдылығын үш арна арқылы күшейте алатынын көрсетті: жергілікті мәдениетті тәжірибе арқылы таныту, туристің дестинация туралы эмоциялық және когнитивтік бейнесін толықтыру, жергілікті қоғамдық тамақтану бизнесін туристік тұтынумен байланыстыру. 2022-2025 жылдардағы материалдар фестивальдік гастрономияның аудиториясы кеңейіп, гастрокартамен байланысы нығайып және кейінгі кезеңде аралас оқиғалық форматтарға енгенін көрсетеді.

Сонымен бірге дестинациялық құндылық қатысушылар санымен ғана өлшенбейді. Түркістан үшін негізгі міндет - жергілікті тағамды тарихи-мәдени брендпен мазмұндық тұрғыдан байланыстыру, фестиваль тәжірибесін тұрақты гастрономиялық маршрутқа жалғастыру және нәтижені туристің нақты мінез-құлқын сипаттайтын көрсеткіштермен өлшеу. Осындай тәсіл гастрономиялық фестивальді бір реттік мерекелік шарадан жергілікті тағамдардың гастрономиялық туризмді дамытудағы рөлін күшейтетін жүйелі дестинациялық құралға айналдырады.

Пайдаланылған әдебиеттер:

1. Leal Londoño M. P., Georgescu-Paquin A., Arcos-Pumarola J. Food festivals as builders of the image of a tourist destination // Journal of Policy Research in Tourism, Leisure and Events. - 2022. - Vol. 14, No. 3. - P. 262-278. - DOI: 10.1080/19407963.2022.2076690.

2. Tran A. T. T., Le A. D., Bui P. L. A., Le V. V., Vu L. T. Food festival research review in contemporary tourism // International Journal of Tourism Cities. - 2023. - Vol. 9, No. 2. - P. 325-347. - DOI: 10.1108/IJTC-01-2022-0009.

3. Huete-Alcocer N., Hernandez-Rojas R. D. Does local cuisine influence the image of a World Heritage destination and subsequent loyalty to that destination? // International Journal of Gastronomy and Food Science. - 2022. - Vol. 27. - Art. 100470. - DOI: 10.1016/j.ijgfs.2022.100470.

4. Hsu F. C., Agyeiwaah E., Chen L. I. L. Examining food festival attendees' existential authenticity and experiential value on affective factors and loyalty: An application of stimulus-organism-response paradigm // Journal of Hospitality and Tourism Management. - 2021. - Vol. 48. - P. 264-274. - DOI: 10.1016/j.jhtm.2021.06.014.

5. Choo H., Park D. B., Petrick J. F. Festival tourists' loyalty: The role of involvement in local food festivals // Journal of Hospitality and Tourism Management. - 2022. - Vol. 50. - P. 57-66. - DOI: 10.1016/j.jhtm.2021.12.002.

6. Pavlović D., Gavrić N., Marić Stanković A., Bratić M., Golubović N., Ivanović M. Gastronomic Festivals as Drivers of Destination Image and Visitor Loyalty: Evidence from the Belmužijada in Serbia // Tourism and Hospitality. - 2026. - Vol. 7, No. 6. - Art. 149. - DOI: 10.3390/tourhosp7060149.

7. Қазақстан Республикасының туристік саласын дамытудың 2023-2029 жылдарға арналған тұжырымдамасын бекіту туралы: Қазақстан Республикасы Үкіметінің 2023 жылғы 28 наурыздағы № 262 қаулысы [Электрондық ресурс]. - URL: https://adilet.zan.kz/rus/docs/P2300000262 (қаралған күні: 06.10.2026).

8. Түркістан облысының туризм басқармасы. Түркістан қаласында алғашқы рет «Gastro Fest Turkistan - 2022» ұйымдастырылып, өткізілді [Электрондық ресурс]. - 2022. - URL: https://www.gov.kz/memleket/entities/turkestan-turizm/press/news/details/458250 (қаралған күні: 06.10.2026).

9. Түркістан облысы әкімінің баспасөз қызметі. Түркістанда Наурыз мейрамында «Gastro Fest Turkistan - 2023» фестивалі өтті [Электрондық ресурс]. - 2023. - URL: https://www.gov.kz/memleket/entities/ontustik/press/news/details/525612 (қаралған күні: 06.10.2026).

10. Түркістан облысының Өңірлік коммуникациялар қызметі. Түркістан - Түркі мемлекеттерінің туристік астанасы атауына орай шаһарда 100-ге жуық іс-шара өтеді [Электрондық ресурс]. - 2024. - URL: https://www.gov.kz/memleket/entities/ontustik/press/news/details/725241 (қаралған күні: 06.10.2026).

11. Түркістан облысы әкімінің баспасөз қызметі. Түркістанда тарихи оқиға - Орта Азияда теңдесі жоқ «Ескек есу» арнасы ашылды [Электрондық ресурс]. - 2025. - URL: https://www.gov.kz/memleket/entities/ontustik/press/news/details/1067535 (қаралған күні: 06.10.2026).

12. Түркістан облысы әкімінің баспасөз қызметі. Түркістан облысы еліміздегі ең танымал туристік өңірлер қатарында [Электрондық ресурс]. - 2026. - URL: https://www.gov.kz/memleket/entities/ontustik/press/news/details/1265453 (қаралған күні: 06.10.2026).

ЕСТЕСТВЕННЫЕ НАУКИ - ЖАРАТЫЛЫСТАНУ ҒЫЛЫМДАРЫ - NATURAL SCIENCES

ӘОЖ 53:001.83

Нургельдина Асель Ермековна
жаратылыстану ғылымдарының магистрі,
математика және физика кафедрасының аға оқытушысы,
Ахмет Байтұрсынұлы атындағы
Қостанай өңірлік университеті
(Қазақстан, Қостанай қ.)

Серикбаева Жанерке Алимжановна
6В05301 - Физика ББББ 2-курс студенті,
машина жасау, энергетика және
ақпараттық технологиялар факультеті,
Ахмет Байтұрсынұлы атындағы
Қостанай өңірлік университеті
(Қазақстан, Қостанай қ.)

ШЕКАРАСЫЗ ФИЗИКА: ӘР ЕЛДЕГІ ФИЗИКАЛЫҚ БІЛІМДІ САЛЫСТЫРУ

Аннотация: Мақалада әлемнің түрлі елдерінде дене білімін қалыптастыру және дамыту ерекшеліктері қаралады. Физикалық білім беру тәсілдері, ғылыми зерттеулердің негізгі бағыттары және физиканың Қазақстанның, Ресейдің, АҚШ-тың, Еуропа елдерінің, Қытай мен Жапонияның технологиялық дамуымен байланысы салыстырылады. Білім беру жүйелеріндегі, ғылыми инфрақұрылымдағы және зерттеу басымдықтарындағы айырмашылықтарға ерекше назар аударылады. Физиканың іргелі заңдары әмбебап сипатқа ие екені, ал халықаралық ғылыми ынтымақтастық білім алмасуға, технологияларды дамытуға және жаһандық ғылыми міндеттерді шешуге ықпал ететіні көрсетілген.

Түйінді сөздер: физика, физикалық білім, ғылыми зерттеулер, STEM-білім, технологиялар, ғылыми ынтымақтастық, халықаралық ғылым, Қазақстан, инновациялар.

Физика ғылыми білім жүйесінде ерекше орын алады, өйткені ол адамның географиялық жағдайына қарамастан бірдей әрекет ететін табиғаттың іргелі заңдылықтарын зерттейді. Механика, электромагнетизм, термодинамика және кванттық физика заңдары кез келген елде қолданылады, алайда оларды зерделеу және практикалық пайдалану тәсілдері әртүрлі. Дене білімінің мазмұны мен зерттеулердің дамуына ғылыми инфрақұрылым деңгейі, мемлекеттік басымдықтар, технологияларға қол жеткізу және ұлттық білім беру жүйелерінің ерекшеліктері әсер етеді. Физиканың қазіргі дамуын жекелеген мемлекеттер шеңберінде ғана қарастыру мүмкін емес. Жаңа материалдар жасау, ғарышты зерттеу, кванттық технологияларды, ядролық энергетика мен медициналық диагностиканы дамыту әртүрлі елдерден ғылыми ресурстар мен мамандарды біріктіруді талап етеді. Сондықтан дене білімі бір мезгілде ұлттық және халықаралық деңгейде қалыптасады. Әртүрлі елдерді салыстыру болашақ физиктерді

даярлау үшін қандай тәсілдер пайдаланылатынын, қандай ғылыми бағыттар барынша дамитынын және іргелі зерттеулер технологиялық шешімдерге қалай айналатынын анықтауға мүмкіндік береді. Қазақстанның осы ғылыми кеңістіктегі жағдайы және оның шетелдік зерттеу орталықтарымен өзара іс-қимылы ерекше қызығушылық тудырады [1, 2].

Мақаланың мақсаты - түрлі елдерде физикалық білімді қалыптастыру, тарату және практикалық қолдану ерекшеліктерін салыстыру және қазіргі заманғы физиканы дамыту үшін халықаралық ынтымақтастықтың маңызын көрсету.

1 Физикалық білімнің әмбебаптығы

Физикалық білім әмбебап сипатқа ие, өйткені зерттеушінің мемлекеттік шекарасына, тіліне немесе ұлтына байланысты емес табиғат заңдылықтарын сипаттайды. Мысалы, Ньютон, энергияны сақтау, электромагниттік өзара әрекеттесу және термодинамика заңдары эксперимент жүргізілетін жерге қарамастан бірдей физикалық жағдайларда қолданылады. Физиканың әмбебаптығы ең алдымен нәтижелердің қайталануына негізделген. Егер әртүрлі елдердің ғалымдары эксперименттің бірдей әдістері мен шарттарын пайдаланса, олар жол берілетін қателік шегінде салыстырмалы нәтижелер алуы тиіс. Осының арқасында ғылыми жаңалықтар тәуелсіз тексеріліп, жалпы ғылыми білімнің бір бөлігіне айналуы мүмкін. Сондай-ақ физиканың дамуы қазіргі жаңалықтар сирек бір ғана елге тиесілі екенін көрсетеді. Ғылыми идеялар әр түрлі мемлекеттердің зерттеушілерінің нәтижелерімен үнемі толықтырылып отырады. Теориялық модельдер, эксперименттік деректер мен техникалық әзірлемелер бірыңғай білім жүйесіне біріктіріледі. Бұл ретте ұлттық айырмашылықтар физикалық заңдардың өзінде емес, оларды зерделеу және қолдану тәсілдерінде көрінеді. Кейбір елдер іргелі теорияға, басқалары – эксперименттік зерттеулерге, инженерлік әзірлемелерге немесе өнеркәсіп пен медицинаға физикалық жаңалықтарды енгізуге баса назар аударады. Физикалық білімнің әмбебаптығы табиғат бірыңғай заңдылықтарға бағынады, ал оларды зерделеу бүкіл әлемнің ғылыми қоғамдастығы үшін ортақ процеске айналады. Бұл физиканы халықаралық ынтымақтастықтың ең жарқын салаларының біріне айналдырады. [1, 3]

2 Дене білімі: ұлттық тәсілдердің айырмашылығы

Түрлі елдерде физикалық білім берудің ортақ негізі бар - механиканы, жылу құбылыстарын, электрді, магнетизмді, оптиканы, атом және ядролық физиканы зерделеу. Алайда бағдарламалардың мазмұны, оқыту әдістері және практикалық сабақтарды ұйымдастыру айтарлықтай өзгеше болуы мүмкін. Бұл ерекшеліктер білім беру дәстүрлерінің, технологиялық даму деңгейінің және экономика қажеттілігінің әсерінен қалыптасады. Қазақстанда физика жаратылыстану-ғылыми даярлықтың бір бөлігі болып табылады және инженерлік және техникалық мамандықтар үшін ерекше маңызды. Білім беру процесінде теориялық сабақтар, міндеттерді шешу, зертханалық жұмыстар және сандық білім беру ресурстары пайдаланылады. Физика математика, информатика және инженерлік пәндермен бірігетін STEM-тәсіл өзекті бағыттардың бірі болып табылады. Ресейде дәстүрлі түрде математикалық дайындыққа, күрделі міндеттерді шешуге және теориялық ойлауды дамытуға көп көңіл бөлінеді. Нақты ғылымдарға қызығушылықпен оқушыларды анықтауға мүмкіндік беретін физика олимпиадалары мен мамандандырылған мектептер маңызды рөл атқарады. Мұндай тәсіл техникалық және физика-математикалық жоғары оқу орындарында одан әрі оқыту үшін берік база қалыптастырады. АҚШ-та физикалық білім беру білімді практикалық қолдану,

зертханалық зерттеулер және жобалау қызметі төңірегінде жиі құрылады. Студенттерге зерттеу мәселелерін өз бетінше тұжырымдауға, эксперименттер жүргізуге және заманауи жабдықтармен жұмыс істеуге мүмкіндік беріледі. Физика инженериямен, компьютерлік технологиялармен және басқа да STEM пәндерімен тығыз байланысты. Еуропа елдерінде пәнаралық білім беру бағдарламалары мен академиялық ұтқырлық кеңінен таралған. Студенттер халықаралық жобаларға қатыса алады, басқа елдерде оқи алады немесе зерттей алады. Университеттік білім берудің ғылыми зертханалармен және зерттеу орталықтарымен байланысының маңызы зор. Қытай мен Жапонияда математикаға, жаратылыстану ғылымдары мен технологиялық даярлыққа көп көңіл бөлінеді. Дене білімі инженерлік бағыттармен, электроникамен, материалтанумен, энергетикамен және қазіргі заманғы технологиялармен тығыз байланысты. Осылайша, ұлттық тәсілдер арасындағы айырмашылықтар ең алдымен теория мен практиканың арақатынасында, оқыту әдістерінде, зертханалық базаға қол жеткізуде және білім берудің ғылыми зерттеулермен байланысында көрінеді. Бұл ретте физиканың іргелі мазмұны барлық елдер үшін ортақ болып қалады, бұл студенттер мен ғалымдарға халықаралық ғылыми ортада оқуды және ынтымақтастықты жалғастыруға мүмкіндік береді [2, 6].

3 Негізгі физикалық зерттеу бағыттарын салыстыру

Әртүрлі елдер арасындағы айырмашылық физика заңдарының өзінде емес, ғылыми зерттеулерде басымдық берілетін бағыттарда байқалады. Әр мемлекет өзінің ғылыми әлеуетіне, технологиялық мүмкіндіктеріне және экономикалық қажеттіліктеріне сәйкес белгілі бір физика салаларын дамытуға көбірек көңіл бөледі [4, 5, 6].

1-кесте. Әлем елдеріндегі физикалық зерттеулердің негізгі бағыттары мен ғылыми басымдықтарын салыстырмалы талдау

Ел	Негізгі зерттеу бағыттары	Ерекшелігі
Қазақстан	Ядролық физика, ғарыштық зерттеулер, материалтану, энергетика	Ядролық зерттеулер мен ғарыштық технологияларды дамытуға, сондай-ақ материалдардың қасиеттерін зерттеуге мән беріледі
Ресей	Ядролық физика, плазма физикасы, астрофизика, теориялық физика	Фундаменталды зерттеулер мен теориялық физика бойынша қалыптасқан ғылыми мектептері бар
АҚШ	Кванттық физика, бөлшектер физикасы, астрофизика, жаңа материалдар	Фундаменталды зерттеулерді жоғары технологиялармен және инженерлік жобалармен байланыстырады
Германия	Кванттық технологиялар, материалтану, энергетика, бөлшектер физикасы	Ғылыми зерттеулерді заманауи зертханалық инфрақұрылыммен және өнеркәсіппен ұштастырады
Ұлыбритания	Кванттық физика, астрофизика, бөлшектер физикасы	Университеттер мен халықаралық ғылыми орталықтардың зерттеулеріне белсенді қатысады
Қытай	Кванттық технологиялар, ядролық физика, материалтану, ғарыш физикасы	Ғылыми инфрақұрылымды кеңейтіп, заманауи технологиялық бағыттарға үлкен көңіл бөледі

Жапония	Материалтану, ядролық физика, кванттық технологиялар, электроника	Фундаменталды физиканы инженерия және жоғары технологиялармен байланыстырады
Үндістан	Астрофизика, ядролық физика, бөлшектер физикасы, ғарыштық зерттеулер	Ғарыштық және ядролық зерттеулермен қатар фундаменталды ғылымды дамытуға көңіл бөледі
Франция	Ядролық физика, бөлшектер физикасы, астрофизика, энергетика	Халықаралық ғылыми жобаларға және ірі зерттеу инфрақұрылымдарына қатысады

Кестеден көрінгендей, мемлекеттердің физикадағы ғылыми басымдықтары әртүрлі болғанымен, олардың арасында ортақ бағыттар көп. Кванттық физика, ядролық физика, материалтану, ғарыштық зерттеулер және энергетика бірнеше елде қатар дамып келеді. Бұл физиканың ұлттық шекаралармен шектелмейтінін көрсетеді. Бір елде жасалған теориялық жаңалық басқа мемлекетте жүргізілетін экспериментке негіз болуы мүмкін, ал халықаралық зерттеу жобалары әртүрлі елдердің ғылыми тәжірибесін біріктіреді [4, 6].

4 Физика және ұлттық технологиялық басымдықтар

Физиканы дамыту мемлекеттің технологиялық және экономикалық мақсаттарымен тікелей байланысты. Елдер ғылыми ресурстарын өздері үшін маңызды салаларға бағыттайды. Энергетикаға басымдық беретін мемлекеттер ядролық энергетика, термоядролық синтез, күн және жел энергетикасы, энергияны сақтау технологияларын зерттейді. Электроникасы дамыған елдер жартылай өткізгіштер, наноматериалдар, фотоника және кванттық технологияларға инвестиция салады. Ғарыш саласы қалыптасқан мемлекеттер астрофизикасы, спутниктік технологиялар, материалдардың ғарыш жағдайындағы қасиеттері және ғарыш аппараттарының физикасын зерттейді. Энергетикада физикалық зерттеулер энергияны өндірудің тиімділігін арттыруға және шығынды азайтуға бағытталады. Фотоэлектрлік материалдар күн сәулесін электр энергиясына айналдыру тиімділігін арттыруға мүмкіндік береді. Аккумуляторлар мен жаңа энергия сақтау жүйелерін жетілдіру үшін электрод материалдарының, иондардың қозғалысының және электрохимиялық процестердің физикасы зерттеледі. Медицинада физика диагностикалық және емдеу технологияларының негізін құрайды. Магниттік-резонанстық томография, рентгендік диагностика, сәулелік терапия және ультрадыбыстық зерттеу физикалық құбылыстарды медициналық мақсатта қолдануға негізделген. Сондықтан медициналық технологияларды дамыту физиктерді, дәрігерлерді, инженерлерді және IT мамандарын бірлесіп жұмыс істеуді талап етеді. Микроэлектроникада жартылай өткізгіштердің қасиеттері, электрондардың қозғалысы, нанокұрылымдар және кванттық құбылыстар зерттеледі. Бұл зерттеулер процессорлар, сенсорлар, байланыс құрылғылары және есептеу жүйелерінің дамуына әсер етеді. Ғарыш технологияларында физика ғарыш аппараттарының қозғалысын есептеу, атмосферадан тыс ортаны зерттеу, радиациялық әсерді бағалау және ғарыштық байланыс жүйелерін жасау үшін қолданылады. Сондықтан физика ұлттық технологиялық саясаттың теориялық негізінің бірі болып саналады. Іргелі зерттеу нәтижесі ұзақ мерзімде энергетика, медицина, көлік, байланыс, электроника және ғарыш салаларында нақты технологияға айналуы мүмкін [6, 7].

5 Қазақстан жаһандық физика кеңістігінде

Қазақстанның физика ғылымындағы әлеуеті ядролық зерттеулермен, ғарыштық технологиялармен, материалтанумен және энергетикамен байланысты. Елдегі ғылыми бағыттардың бір бөлігі Қазақстанның табиғи ресурстарына, ғылыми инфрақұрылымына және тарихи қалыптасқан зерттеу базасына сәйкес дамып келеді. Ядролық физика Қазақстанның маңызды ғылыми бағыты болып табылады. Семей ядролық полигонының тарихы елде радиациялық қауіпсіздік, ядролық материалдар және радиацияның қоршаған ортаға әсерін зерттеудің қажеттілігін қалыптастырды. Қазақстанда Ұлттық ядролық орталық ядролық физика, радиациялық материалтану, термоядролық зерттеулер және ядролық технологиялар бойынша ғылыми жұмыстар жүргізетін негізгі ұйымдардың бірі. Ғарыш физикасы Қазақстан үшін географиялық және технологиялық жағынан маңызды бағыт болып табылады. Байқоңыр ғарыш айлағы Қазақстан аумағында орналасқан және ғарыштық зерттеулердің халықаралық тарихында ерекше орын алады. Ғарыштық технологияларды дамыту жерді қашықтықтан зондтау, спутниктік байланыс, атмосфераны бақылау және ғарыштық аппараттардың жұмыс істеу физикасымен байланысты. Материалтану саласында металдар, қорытпалар, керамикалық материалдар және басқа да құрылымдардың физикалық қасиеттері зерттеледі. Мұндай зерттеулер энергетикада, өнеркәсіпте, құрылыста және машина жасауда қолданылатын материалдардың беріктігін, жылуға төзімділігін және пайдалану мерзімін арттыруға мүмкіндік береді. Энергетика физикасы да Қазақстан үшін маңызды. Елдің энергетикалық жүйесінде дәстүрлі энергия көздерімен қатар жаңартылатын энергетиканы дамыту мәселесі қаралады. Күн және жел энергетикасының тиімділігін арттыру климаттық жағдайларды, материалдардың қасиеттерін және энергияны сақтау технологияларын зерттеуді қажет етеді. Қазақстанның әлемдік физика кеңістігіне қатысуының маңызды бөлігі - халықаралық ғылыми ынтымақтастық. Бірлескен зерттеулер қазақстандық ғалымдарға шетелдік зертханалардың тәжірибесін пайдалануға, ғылыми жабдықтарға қол жеткізуге және халықаралық жобаларға қатысуға мүмкіндік береді. Студенттер мен жас зерттеушілер үшін академиялық мобильдік және халықаралық ғылыми жобалар кәсіби тәжірибе жинаудың маңызды тетігі болып табылады. Қазақстанның физика ғылымын дамыту үшін зертханалық инфрақұрылымды жаңарту, жас ғалымдарды қолдау, халықаралық жобалар санын ұлғайту және университет пен ғылыми-зерттеу орталықтарының байланысын күшейту маңызды. Бұл бағыттар отандық физикалық зерттеулерді қазіргі әлемдік ғылыми үрдістермен байланыстыруға мүмкіндік береді. [6, 8]

6 Халықаралық ынтымақтастық – қазіргі физиканың негізі

Қазіргі физикадағы көптеген зерттеулер бір мемлекеттің ғылыми мүмкіндіктерімен шектелмейді. Үлкен ғылыми жобаларға күрделі жабдықтар, жоғары қаржыландыру, әртүрлі мамандар және ұзақ мерзімді зерттеу бағдарламалары қажет. Сондықтан халықаралық ынтымақтастық физикалық зерттеулердің маңызды ұйымдастырушылық негізіне айналды. Халықаралық ынтымақтастықтың нақты мысалы — CERN. Мұнда әртүрлі мемлекеттердің ғалымдары бөлшектер физикасы бойынша бірлескен эксперименттер жүргізеді. Үлкен адрон коллайдері сияқты күрделі қондырғылардың жұмысына физиктермен қатар инженерлер, бағдарламашылар және деректерді талдау мамандары қатысады. Бірлескен жұмыс эксперимент нәтижелерін үлкен ғылыми қауымдастықтың тексеруіне мүмкіндік береді. Ғарыштық зерттеулерде де халықаралық серіктестік маңызды орын алады. Ғарыш аппараттарын жасау, ғылыми аспаптарды әзірлеу және алынған деректерді өңдеу бірнеше елдің зерттеу топтарының қатысуымен

жүзеге асуы мүмкін. Мұндай жобалар ғарыш физикасы, астрофизика және Жерді зерттеу салаларындағы мәліметтерді біріктіреді. Халықаралық ғылыми байланыстардың тағы бір маңызды бөлігі — ғылыми жарияланымдар мен деректер алмасу. Зерттеушілер өз нәтижелерін халықаралық ғылыми журналдарда жариялап, басқа мамандардың оларды талдауына және қайта тексеруіне мүмкіндік береді. Бұл ғылыми қателерді анықтауға, жаңа гипотезалар ұсынуға және бұрынғы нәтижелерді басқа жағдайларда қолдануға жағдай жасайды. Сонымен қатар халықаралық ынтымақтастық ғылыми білімдегі теңсіздіктерді азайтуға көмектеседі. Зертханалық мүмкіндігі шектеулі елдердің ғалымдары халықаралық жобалар арқылы заманауи әдістермен, ғылыми құралдармен және кәсіби тәжірибемен таныса алады. Осы тұрғыдан алғанда, халықаралық байланыс физиканың дамуын жеделдететін маңызды фактор болып табылады. [4, 5, 6]

7 Әртүрлі елдердің физиктерін не біріктіреді?

Әртүрлі елдерде білім алған физиктердің зерттеу әдістері мен ғылыми дәстүрлері өзгеше болуы мүмкін. Бірақ олардың кәсіби қызметін біріктіретін негізгі қағидалар ортақ. Біріншіден, ғылыми дәлелділік. Физикалық тұжырым бақылау, өлшеу немесе эксперимент арқылы негізделуі керек. Жеке пікір ғылыми дәлелдің орнын баса алмайды. Екіншіден, математикалық сипаттау. Физикалық құбылыстар формулалар, теңдеулер және математикалық модельдер арқылы нақты сипатталады. Бұл әртүрлі елдердің зерттеушілеріне бір ғылыми тілде жұмыс істеуге мүмкіндік береді. Үшіншіден, эксперимент нәтижелерін тексеру. Бір зерттеушінің нәтижесі басқа зерттеу тобымен қайта тексерілуі мүмкін. Нәтижелердің қайталануы ғылыми білімнің сенімділігін арттырады. Төртіншіден, ғылыми терминология мен өлшем бірліктерінің ортақтығы. Халықаралық бірліктер жүйесі (SI) физикалық шамаларды бірдей өлшемдермен сипаттауға мүмкіндік береді. Мысалы, ұзындық метрмен, уақыт секундпен, масса килограммен өлшенеді. Бесіншіден, зерттеу нәтижелерімен алмасу. Ғалымдар мақалалар, конференциялар, ғылыми деректер және халықаралық жобалар арқылы жаңа ақпарат таратады. Бір елде алынған нәтиже басқа мемлекеттегі зерттеудің бастапқы нүктесіне айналуы мүмкін. Осы ортақ қағидалар физиктерді ұлттық шекаралардан тыс біріктіреді. Физиканың ортақ тілі — табиғат заңдары, математикалық модельдер, эксперимент және дәлел. Сондықтан әр елдің ғылыми жүйесі әртүрлі болғанымен, физикалық білімді қалыптастырудың негізгі ғылыми принциптері бүкіл әлемде өзара байланысты [3, 5].

Қорытынды

Физикалық білім ұлттық шекаралармен шектелмейді, себебі табиғат заңдары барлық адамзат үшін ортақ. Әртүрлі мемлекеттерде физиканы оқыту әдістері, зерттеу инфрақұрылымы және ғылыми басымдықтар өзгеше болғанымен, олардың барлығы эксперимент, математикалық модельдеу және ғылыми дәлелге сүйенеді. Қазақстан, АҚШ, Ресей, Еуропа елдері, Қытай және Жапония мысалдары физика ғылымының дамуында әртүрлі бағыттардың қалыптасқанын көрсетеді. Бір мемлекет ядролық немесе ғарыштық зерттеулерге, екіншісі кванттық технологияларға, материалтануға немесе энергетикаға көбірек көңіл бөлуі мүмкін. Мұндай айырмашылықтар ғылыми бәсекелестікті ғана емес, өзара толықтыру мүмкіндігін де қалыптастырады. Қазіргі күрделі ғылыми мәселелерді шешуде халықаралық ынтымақтастықтың маңызы артып келеді. Ортақ зертханалар, ғылыми жарияланымдар, деректер алмасу және халықаралық жобалар физикалық білімнің жылдам таралуына мүмкіндік береді. Қазақстан үшін мұндай байланыстар ғылыми инфрақұрылымды жетілдіру, жас зерттеушілердің тәжірибесін кеңейту және әлемдік ғылыми жобаларға қатысу тұрғысынан маңызды.

Физиканың басты ерекшелігі — оның ортақ ғылыми тіл қалыптастыруында. Елдер арасындағы білім беру және зерттеу тәсілдері әртүрлі болғанымен, физикалық заңдылықтарды түсіндірудің негізі бір. Сондықтан физика ғылымы адамзаттың ортақ білім жүйесін қалыптастыратын және технологиялық дамуға тікелей әсер ететін әмбебап ғылым болып қала береді.

Пайдаланылған әдебиеттер:

1. Halliday D., Resnick R., Walker J. Fundamentals of Physics. — 12th ed. — Hoboken: Wiley, 2021.

2. Serway R. A., Jewett J. W. Physics for Scientists and Engineers. — 10th ed. — Boston: Cengage Learning, 2019.

3. Feynman R. P., Leighton R. B., Sands M. The Feynman Lectures on Physics. — Pasadena: California Institute of Technology, 2010.

4. European Organization for Nuclear Research (CERN). The Large Hadron Collider. — Geneva: CERN. CERN

5. International Union of Pure and Applied Physics (IUPAP). About IUPAP. — IUPAP. IUPAP

6. UNESCO. UNESCO Science Report: The Race Against Time for Smarter Development. — Paris: UNESCO, 2021. UNESCO Science Report

7. National Aeronautics and Space Administration (NASA). NASA Science. — Washington, D.C.: NASA. NASA Science

8. Ұлттық ядролық орталық Қазақстан Республикасы. Ғылыми-зерттеу қызметі. — Курчатов: ҚР ҰЯО. Қазақстан Республикасының Ұлттық ядролық орталығы.

Электронный научный журнал «Central Asian Scientific Journal»

Редактор: Байдильдинов Т.Ж.
Комп.верстка: Хусаинов Е.М.

Электронный научный журнал «Central Asian Scientific Journal»
-2026-34(32)-Астана-ИП DOC
Зарегистрировано и выдано свидетельство
Министерством Информации и Общественного Развития РК
№KZ77VPY00147053 от 17.04.2026 г.
ISSN: 3135-3061

За достоверность публикуемой информации, цитат и иных изложений ответственность несет автор

© Central Asian Scientific Journal, 2026

