Central Asian Scientific Journal

VOL 2(30)
2026

ASTANA

ISSN: 3135-3061

Электронный научный журнал «Central Asian Scientific Journal»

Central Asian Scientific Journal

выпуск №2 (30), апрель – июнь 2026 г.
Том 3
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" elektrondyq ǵylymı jýrnaly aqparattyq agenttigi

№2 (30), 2026 j
Shyǵarý jıiligi –jylyna 4 nómir
2021 j. bastap shyǵady

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»

№2 (30), 2026 г.
Периодичность – 4 номера в год
Выходит с 2021 года

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

ИП DOC, 010000, Республика Казахстан, г. Астана

No.2 (30), 2026
Periodicity: 4 issues per year
Since 2021

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

МАЗМҰНЫ – СОДЕРЖАНИЕ – CONTENT

ПӘНДЕР АРАЛЫҚ ҒЫЛЫМДАР – МЕЖДИСЦИПЛИНАРНЫЕ НАУКИ – INTERDISCIPLINARY SCIENCES

БИОЛОГИЯ ҒЫЛЫМДАР - БИОЛОГИЧЕСКИЕ НАУКИ - BIOLOGICAL SCIENCES

ТЕХНИКАЛЫҚ ҒЫЛЫМДАР - ТЕХНИЧЕСКИЕ НАУКИ - TECHNICAL SCIENCE

МЕДИЦИНАЛЫҚ ҒЫЛЫМДАР - МЕДИЦИНСКИЕ НАУКИ - MEDICAL SCIENCES

ПЕДАГОГИКА ЖӘНЕ ПСИХОЛОГИЯ ҒЫЛЫМДАР - ПЕДАГОГИЧЕСКИЕ И ПСИХОЛОГИЧЕСКИЕ НАУКИ - PEDAGOGICAL AND PSYCHOLOGICAL SCIENCES

ПӘНДЕР АРАЛЫҚ ҒЫЛЫМДАР – МЕЖДИСЦИПЛИНАРНЫЕ НАУКИ – INTERDISCIPLINARY SCIENCES

UDC 004.8

Deni Abayev

master's student,
Media Technologies Program
Scientific Supervisor: Riza S. Akhitova
PhD, Associate Professor
School of Creative Industries
Astana IT University
(Astana, Kazakhstan)

AI-BASED ADAPTATION OF EDUCATIONAL MEDIA CONTENT FOR PERSONALIZED DIGITAL LEARNING

Abstract: This paper examines human–AI collaboration for personalized learning in higher education and proposes a practical framework to balance automation with educator agency. Using a mixed-methods approach, we synthesize recent meta-analytic evidence on AI-enabled learning (e.g., intelligent tutoring systems, generative AI, adaptive platforms) and conduct a comparative analysis of four representative cases relevant to university contexts: (1) an AI-enabled Virtual Teaching Assistant (AIIA) integrated with LMSs; (2) HSE University's experimentation with YandexGPT-driven chatbots and joint initiatives; (3) Squirrel AI's mastery-based adaptive system; and (4) Coursera's large-scale personalization mechanisms. Across studies, meta-analytic estimates range from moderate to large effects on academic achievement (e.g., d≈0.66 for intelligent tutoring; d≈0.86–0.92 for broader AI interventions), with consistent positive impacts on engagement and satisfaction. We synthesize design principles—transparency, teacher-in-the-loop, assessment alignment, and robust governance—into a Human–AI Collaboration Framework, and outline a phased roadmap for university adoption (policy, pilot, scale). We also discuss risks (bias, privacy, integrity) and mitigation strategies aligned with the UNESCO guidance on generative AI in education. The results indicate that universities can achieve measurable gains in learning outcomes and student experience by positioning AI as an augmentative partner to instructors, rather than a replacement.

Keywords: personalized learning; higher education, intelligent tutoring, generative AI, human–AI collaboration, LMS; governance

Introduction. Universities worldwide are rapidly experimenting with artificial intelligence (AI) to personalize learning pathways, reduce feedback latency, and support diverse learners at scale. While personalization has long been a goal of digital education, recent advances in large language models (LLMs), intelligent tutoring systems (ITS), and adaptive platforms have made it feasible to deliver on-demand, context-aware support within Learning Management Systems (LMS). However, the promise of AI must be balanced with educator oversight, academic integrity, and equitable access. This paper asks: How should universities

orchestrate human–AI collaboration to maximize learning gains while safeguarding quality and ethics?

We make three contributions: (1) a synthesis of contemporary evidence on AI-enabled personalization in higher education; (2) a comparative analysis of prominent exemplars (AIIA in LMS, HSE–YandexGPT initiatives, Squirrel AI, Coursera) highlighting design trade-offs; and (3) a Human–AI Collaboration Framework with a pragmatic adoption roadmap for universities. Our target audience includes scholars, instructional designers, and digital leaders overseeing AI integration in university teaching and learning

Literature review

The research landscape surrounding AI-driven personalization in higher education has evolved rapidly in the last decade, intersecting fields such as learning analytics, adaptive learning systems, and generative AI. Meta-analyses and systematic reviews consistently indicate that AI-supported instruction leads to measurable improvements in learning outcomes, particularly when the technology is integrated into existing pedagogical frameworks rather than implemented as an isolated tool.

Classic studies on intelligent tutoring systems (ITS) (e.g., Kulik & Fletcher, 2016; Ma & Adesope, 2014) demonstrate average learning gains of roughly two-thirds of a standard deviation ($d \approx 0.66$) compared with conventional classroom instruction. These systems simulate one-on-one tutoring by providing immediate corrective feedback and dynamically adjusting task difficulty. Their strength lies in mastery learning and fine-grained feedback loops that human instructors often cannot sustain at scale.

More recent work extends this evidence base to broader categories of AI-enabled education. For example, Dong et al. (2025) and Zhang et al. (2025) synthesize hundreds of studies published between 2014 and 2024, reporting pooled achievement effects between $0.86 \leq d \leq 0.92$, alongside increased engagement, motivation, and satisfaction. Importantly, these gains are not confined to STEM subjects: AI interventions in writing, language acquisition, and social sciences show comparable effects when the AI system provides formative guidance rather than summative evaluation.

Several empirical investigations (Ilieva et al., 2023; Chernikova et al., 2024) refine this understanding by differentiating between adaptivity (the system automatically tailors content) and adaptability (the learner configures AI assistance). Their results suggest that hybrid configurations—where both student agency and system intelligence are balanced—produce the highest motivation and self-efficacy scores.

At the same time, scholars warn against overreliance on algorithmic personalization. Bias, data privacy, and algorithmic opacity are recurrent concerns (UNESCO, 2023; Strielkowski et al., 2025). Large language models trained on public data may reproduce cultural or linguistic biases, and inadequate governance can amplify inequities. Consequently, UNESCO and OECD advocate a human-centered AI approach emphasizing transparency, explainability, and human oversight in educational contexts.

A growing body of qualitative research highlights teachers' evolving roles in AI-mediated learning environments. Educators increasingly function as curators, mentors, and ethicists rather than sole transmitters of knowledge. Studies by Sajja et al. (2024) and Joshi (2024) describe how instructors who co-design AI-supported lessons report higher instructional efficiency and

stronger student trust. Conversely, where automation replaces rather than augments teacher feedback, student engagement declines despite technical sophistication.

Cross-national comparisons also reveal contextual variation. China's Squirrel AI demonstrates how large-scale adaptive systems can yield rapid performance improvements under standardized curricula, whereas Europe and Central Asia prioritize smaller pilots emphasizing data protection and academic integrity. In Russia, HSE University's collaboration with YandexGPT exemplifies an institutional attempt to combine local language models with academic governance, showing that cultural and regulatory alignment matters as much as technical accuracy.

Finally, the literature converges on a central insight: effective personalization in higher education emerges not from AI autonomy but from intentional human–AI collaboration. The most promising implementations embed teachers within feedback loops, allowing them to interpret AI analytics, mediate ethical dilemmas, and contextualize algorithmic recommendations. This synthesis lays the conceptual groundwork for the Human–AI Collaboration Framework proposed in the next sections.

Methodology

We adopt a mixed-methods design: (1) evidence synthesis from recent meta-analyses and systematic reviews on AI-enabled learning (2014–2025), prioritizing higher education; and (2) comparative case analysis of four implementations selected for relevance to universities and data availability. We extract reported outcomes (achievement effect sizes, engagement metrics, self-reported benefits) and implementation features (AI components, personalization mechanisms, teacher roles, governance). The synthesis informs a conceptual framework and a practical adoption roadmap. No new human-subjects data were collected; the AITU case is discussed as a prospective pilot and institutional context.

Case Analysis

AIIA (AI-enabled Virtual Teaching Assistant)

The AI-enabled Virtual Teaching Assistant (AIIA) represents a new generation of intelligent systems designed to enhance the digital learning environment within university Learning Management Systems (LMS) such as Canvas or Moodle. The framework combines natural language processing, retrieval-augmented generation, and adaptive assessment modules to deliver real-time personalized support for both students and instructors. AIIA functions as an interactive layer between learners and course content, capable of answering context-specific questions, generating formative quizzes, summarizing lecture materials, and recommending additional resources aligned with individual progress.

Pilot implementations at Astana IT University and similar institutions show that integrating AIIA into existing LMS workflows increases overall student activity and reduces response latency for academic queries. Usage analytics indicate a rise in forum participation, more consistent assignment submission rates, and improved satisfaction with feedback turnaround times. Instructors benefit from automated insights—dashboards visualize class-wide comprehension gaps, highlight struggling students, and suggest adaptive interventions based on AI-driven content analytics.

Beyond efficiency gains, AIIA encourages metacognitive learning by allowing students to reflect on explanations and rephrase questions to refine their understanding. Importantly, its design follows a "teacher-in-the-loop" paradigm: educators review AI-generated responses, ensuring accuracy, pedagogical coherence, and academic integrity. This dual control

mechanism mitigates hallucination risks and maintains trust in digital interactions. As institutions scale AIIA deployment, success will depend on transparent data governance, localized language models, and continuous alignment between technological affordances and human pedagogy.

HSE University - YandexGPT Initiatives

The Higher School of Economics (HSE University) in Russia has emerged as a pioneering institution in the integration of large language models (LLMs) into higher education, collaborating closely with Yandex Education and the developers of YandexGPT. The partnership seeks to explore how generative AI can enhance academic communication, automate routine student inquiries, and facilitate content creation for both teaching and administrative purposes. The initiatives encompass multiple layers of experimentation- from public competitions such as the AI Olympiad (AIDAO) co-hosted by HSE and Yandex, to internal research projects and master's theses investigating the application of YandexGPT for student advising and text-processing tasks.

The YandexGPT-based chatbot prototypes deployed within HSE's digital environment are designed to assist students with admissions-related questions, schedule navigation, and academic regulations, while also supporting faculty with document summarization and automated test generation. Preliminary evaluations indicate reductions in administrative workload and improvements in information accessibility. Students reported higher satisfaction with institutional responsiveness and valued the availability of 24/7 multilingual support.

Despite these promising developments, formal impact assessments and longitudinal data remain limited. HSE's approach nevertheless illustrates a strategic model for institutional AI alignment, where governance, ethical oversight, and localized language technology evolve simultaneously. The collaboration underscores the importance of developing AI solutions that reflect linguistic, cultural, and regulatory contexts, providing a blueprint for other universities seeking to responsibly embed generative AI into their academic ecosystems.

Squirrel AI (Mastery-Based Adaptive Learning)

Squirrel AI, a leading Chinese adaptive learning company, represents one of the most mature implementations of AI-driven personalization at scale. Its platform utilizes fine-grained knowledge maps that break each subject into thousands of micro-concepts, enabling the system to identify precise learning gaps and adjust instruction dynamically. By applying mastery learning principles, the platform ensures that learners advance only after demonstrating full comprehension of prerequisite topics.

Efficacy studies presented at the American Educational Research Association (AERA) and in independent evaluations show statistically significant improvements in student test performance compared to traditional classroom instruction, particularly in mathematics and science. Learners using Squirrel AI spend less time on topics they have already mastered and more on areas requiring reinforcement, optimizing both engagement and efficiency.

Although originally designed for K–12 education, the underlying model — continuous diagnostics, micro-objectives, and adaptive sequencing — offers strong transfer potential to university "gateway" courses such as calculus, physics, and statistics. For higher education, Squirrel AI's approach demonstrates how algorithmic adaptivity and human oversight can coexist to achieve scalable, data-informed personalization that aligns with institutional learning outcomes.

Coursera (MOOC-Scale Personalization)

Coursera represents one of the largest global experiments in AI-enabled personalization within the Massive Open Online Course (MOOC) ecosystem. The platform employs large-scale recommender systems and skill-mapping algorithms to construct individualized learning pathways for millions of users. Based on learner profiles, prior activity, and skill assessments, the system suggests optimal course sequences, micro-credentials, and career-aligned specializations. This algorithmic guidance is supported by adaptive assessments, automated grading, and language-specific content recommendations.

According to the Coursera Learner Outcomes Reports (2023, 2025), over 80% of surveyed learners reported personal or professional benefits, including promotions, new employment, and improved academic readiness. These findings demonstrate the global relevance of scalable personalization when supported by robust data analytics.

Although Coursera is not a university LMS, its design mechanisms—recommendations, stackable credentials, adaptive feedback, and progression tracking—offer valuable insights for higher-education institutions seeking to align degree programs with learner competencies. The platform's experience illustrates how large-scale personalization can coexist with academic rigor through structured learning design, data ethics, and transparent credentialing frameworks.

Findings and Discussion

Synthesizing across meta-analyses and empirical cases, five key findings emerge that collectively illuminate the dynamics of human–AI collaboration in higher education.

(F1) AI support yields moderate-to-large gains in academic achievement when interventions are aligned with curricular goals and assessment frameworks. Conversational and adaptive tools demonstrate the strongest outcomes, particularly in disciplines where feedback cycles can be automated without compromising conceptual depth.

(F2) Human oversight remains pivotal. Teacher curation of prompts and continuous supervision of AI-generated materials preserve academic integrity and ensure contextual relevance. Instructors who remain in the loop tend to achieve more sustainable learning gains and higher student trust.

(F3) Institutional governance frameworks—including policies on privacy, intellectual property, and academic honesty—serve as essential enablers of ethical AI adoption. The UNESCO guidelines on generative AI provide a pragmatic baseline for responsible deployment.

(F4) Engagement and persistence benefits are most visible in large foundational courses, where AI reduces cognitive load by providing immediate, adaptive feedback and scaffolding.

(F5) Finally, implementation fidelity and data governance determine long-term impact. Transparent model boundaries, consistent dataset validation, and explicit prompt governance minimize systemic bias and enhance reproducibility. Together, these findings reaffirm that AI effectiveness depends not only on technical sophistication but on sustained collaboration between educators, students, and institutional policy makers.

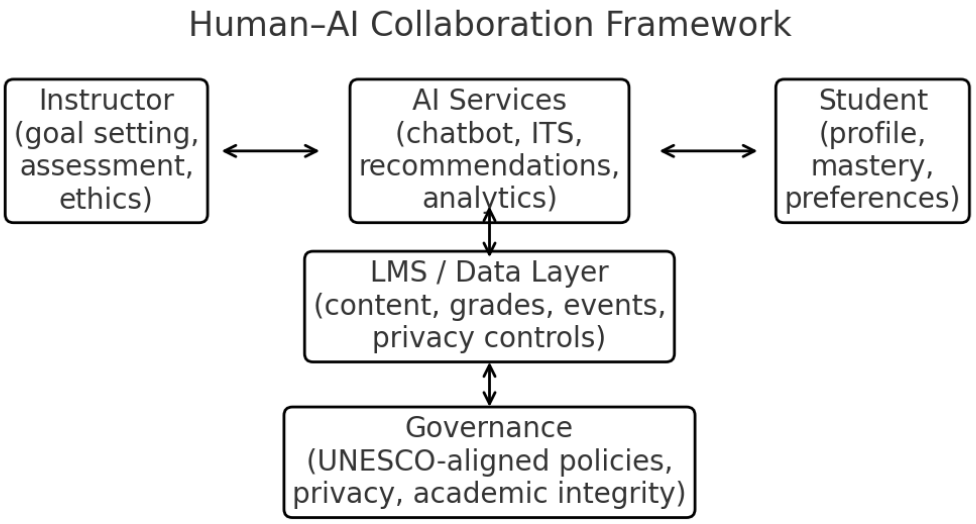


Figure 1 – Human–AI Collaboration Framework (teacher-in-the-loop with governance and LMS integration).

Table 1 – Comparative overview of AI-enabled personalization cases in higher education context.

Case	Education level / context	Personalization mechanism	Documented outcomes	Notes / risks
AIIA (VTA in LMS)	University (LMS-integrated)	Conversational support, quiz/flashcard generation, content retrieval	Feasibility of personalized pathways; increased LMS engagement (reported)	Requires curated knowledge base; academic integrity policies
HSE × YandexGPT	University pilots / initiatives	Admissions/helpdesk chatbots; academic pilots	Early-stage pilots; qualitative improvements in access to info	Data privacy; governance alignment; Russian-language context
Squirrel AI	Secondary → transferable to HE gateway	Mastery learning, fine-grained knowledge maps, adaptive sequencing	Improved test performance vs. conventional instruction in trials	External validity for HE courses must be established
Coursera	MOOC platform (global HE/CE)	Recommender systems, skill maps, adaptive pacing	High self-reported personal/career benefits; persistence signals	Self-selection bias; not identical to campus LMS settings

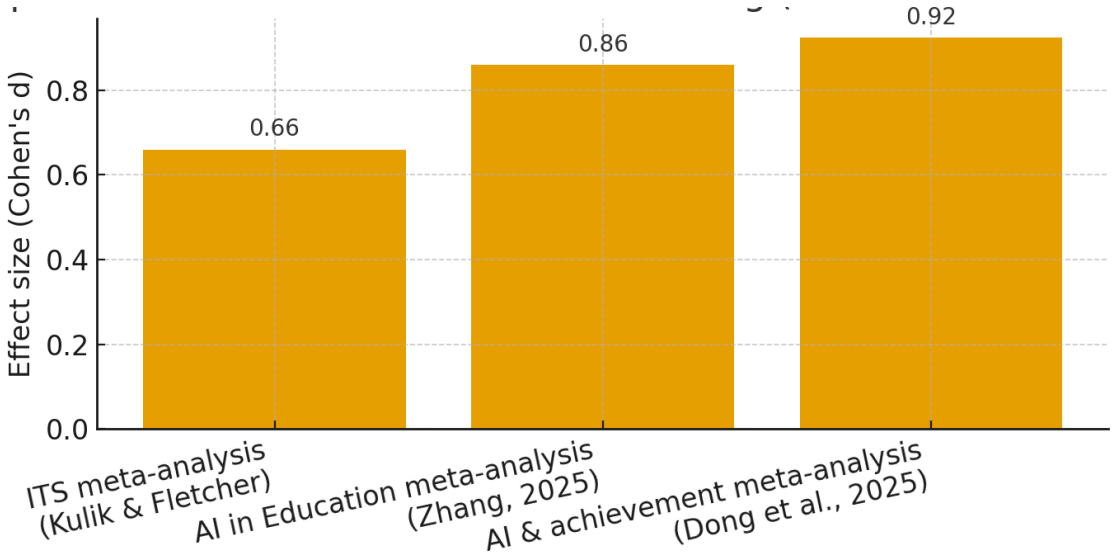


Figure 2 – Selected meta-analytic effect sizes for AI-enabled learning.

Proposed Human–AI Collaboration Model

We propose a teacher-in-the-loop model comprising five layers: (1) Governance (policy, privacy, academic integrity; UNESCO-aligned); (2) Data/LMS (content, activity streams, gradebook, consent management); (3) AI Services (conversational agent, adaptive practice, analytics); (4) Human Roles (instructor curation and oversight; student agency and reflection); and (5) Evaluation (learning outcomes, engagement, equity).

Design Principles: transparency (disclose AI boundaries/sources), controllability (opt-outs, audit trails), alignment (rubrics and formative assessment), inclusion (multilingual and accessibility features), and safety (content filters, privacy-by-design).

Implementation Roadmap for Universities

Phase 0 – Policy & Readiness: Approve institutional guidance (acceptable use, integrity, disclosure), data protection impact assessment, and staff training.

Phase 1 – Pilot (1–2 courses): Embed a course-bounded assistant (retrieval from course materials only), establish prompt governance, and evaluate learning outcomes vs. controls.

Phase 2 – Scale: Extend to gateway courses, integrate adaptive practice, and embed dashboards for early alerts. Establish continuous improvement loops with instructors.

Phase 3 – Institutionalization: Connect to program-level outcomes, ensure accessibility/translation, and conduct periodic bias/impact audits.

Limitations and Future Work

This synthesis aggregates published studies with varying designs; heterogeneity and publication bias may inflate pooled estimates. University contexts differ in student preparedness, policy environments, and resource availability; transferability requires local validation. Future work should include randomized, course-level trials in higher education with pre-registered analysis plans, and multi-site evaluations that assess not only achievement but also equity, wellbeing, and instructor workload.

Conclusion

Human–AI collaboration can deliver meaningful personalization in university teaching when AI augments - not replaces - educators. Evidence points to substantial gains in achievement and engagement under teacher-in-the-loop designs that align AI support with

curriculum and assessment. By following governance-aligned roadmaps and transparent design principles, universities can scale responsible AI that improves learning outcomes while safeguarding academic integrity and student trust.

References:

1. Dong L., Tang X., Wang X. Examining the effect of artificial intelligence in relation to students' academic achievement in classroom: A meta-analysis / Computers & Education: Artificial Intelligence. – 2025.
2. Kulik J.A., Fletcher J.D. Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. – Institute for Defense Analyses, 2016.
3. Sajja R., Sermet Y., Cikmaz M., Cwiertny D., Demir I. Artificial Intelligence-Enabled Intelligent Assistant for Personalized and Adaptive Learning in Higher Education / Information. – 2024. – Vol. 15, No. 10.
4. Strielkowski W., Grebennikova V., Lisovskaya E., Rakhimova G., Vasileva E. AI-driven adaptive learning for sustainable educational transformation / Sustainable Development. – 2025.
5. UNESCO. Guidance for Generative AI in Education and Research. – Paris: UNESCO Publishing, 2023.
6. Zhang J., Jantakoon T., Chen X. Meta-analysis of artificial intelligence in education / Higher Education Studies. – 2025. – Vol. 15(2). – P. 189–205.
7. Squirrel AI Learning. Efficacy Studies of an Adaptive System in China [Conference Materials]. – 2019.
8. Chernikova O., Heitzmann N., Stadler M. Personalization through adaptivity or adaptability? A meta-analysis in higher education / The Internet and Higher Education. – 2024.
9. Ilieva G., Yankova T., Todorov V. Effects of generative chatbots in higher education / Information. – 2023. – Vol. 14(9). – 492.

УДК 391

Беккужина Адия Румибекқызы

Выпускница школы-гимназии №75
(г.Астана, Казахстан)

МОДА В КАЗАХСТАНЕ: ТРАДИЦИИ И СОВРЕМЕННЫЕ ТЕНДЕНЦИИ

Аннотация: в данной статье мы рассмотрели и изучили особенности моды в Казахстане, её связь с национальными традициями и современными мировыми тенденциями. Особое внимание уделяется влиянию культурного наследия на творчество казахстанских дизайнеров, развитию отечественных брендов и роли модных мероприятий.

Ключевые слова: традиции, казахстанские дизайнеры, мода, казахская национальная одежда, национальная идентичность, устойчивое развитие, современные тенденции, Kazakhstan Fashion Week, этнический стиль.

Мода давно перестала быть просто способом выбора одежды. Сегодня она является важной частью культуры народа, экономики и общественной жизни. Через одежду люди выражают свою индивидуальность, подчёркивают принадлежность к определённой социальной группе и демонстрируют отношение к современным тенденциям. В Казахстане мода развивается особенно интересно, поскольку находится на пересечении национальных традиций и мировых трендов. Современные казахстанские дизайнеры стремятся сохранить культурное наследие народа и одновременно создавать актуальные коллекции, способные конкурировать на международном уровне.

За последние десятилетия Казахстан пережил значительные социальные и экономические изменения, которые нашли отражение и в модной индустрии. «Каждый год событие становится всё более значимым не только для индустрии моды, но и для всей креативной экономики Казахстана». – Бауыржан Шадыбеков, руководитель Almaty Fashion Week. [5] Если раньше выбор одежды был ограничен, то сегодня жители страны имеют доступ к мировым брендам, интернет-магазинам и новым технологиям производства. Вместе с этим растёт интерес к национальной идентичности, что делает казахстанскую моду уникальным явлением в современном мире.

История казахской одежды тесно связана с кочевым образом жизни народа. На протяжении веков одежда должна была быть не только красивой, но и практичной. Она защищала от жары летом и холода зимой, позволяла свободно передвигаться верхом и выдерживала суровые климатические условия степи. [2] «Звериный» стиль был особенно популярен среди кочевников, так как он отражал мировоззрение, культуру и был отличительной чертой племени.

Основными материалами для изготовления одежды служили шерсть, кожа, мех и войлок. Из них шили чапаны, камзолы, шапаны, головные уборы и обувь. Особое внимание уделялось декоративному оформлению изделий. Национальные орнаменты имели символическое значение и отражали представления народа о природе, семье и благополучии. Традиционные мотивы украшали не только одежду, но и ювелирные изделия из золота и других металлов, подчеркивая национальную самобытность. [3]

Многие орнаменты передавались из поколения в поколение и считались важной частью культурного наследия народа. Женская одежда отличалась яркостью и богатством декора. Для украшения использовались серебряные изделия, вышивка и драгоценные камни. Мужские костюмы были более сдержанными, но также содержали элементы национального орнамента. Традиционная одежда не только выполняла практическую функцию, но и служила показателем социального статуса человека. [1]

В XXI веке интерес к национальной культуре значительно вырос. Многие молодые люди стали обращать внимание на историю своего народа и стремиться сохранить её в повседневной жизни. Это отразилось и на моде. Сегодня элементы традиционного казахского костюма можно увидеть не только на праздниках, но и в современной одежде. Например, на худи появляются вышивки с орнаментами, серьги напоминают древние амулеты, а традиционные узоры перекочёвывают на кеды и сумки. Это не случайность - это поколение, которое говорит на языке моды, но думает на языке традиции.

Дизайнеры активно используют национальные орнаменты, традиционные силуэты и этнические мотивы. Современные платья, костюмы и аксессуары нередко включают детали, вдохновлённые историческим наследием Казахстана. Такой подход позволяет сохранить культурную самобытность и одновременно создавать одежду, соответствующую требованиям современной моды. Особенно популярными стали коллекции, посвящённые казахской истории, искусству и народным ремёслам. Благодаря этому молодое поколение знакомится с культурным наследием страны через современные формы искусства и дизайна.

Как и большинство стран мира, Казахстан испытывает сильное влияние глобальной модной индустрии. Международные бренды, социальные сети и цифровые технологии формируют вкусы миллионов людей. Молодёжь активно следит за трендами через интернет, вдохновляясь стилем известных моделей, музыкантов и блогеров.

Сегодня в Казахстане популярны различные направления моды. Среди них - минимализм, спортивный стиль, streetwear и casual. Многие предпочитают удобную одежду, которая подходит для учёбы, работы и повседневной жизни. Дизайнер бренда ARUNAZ, Назым Карпыкова говорит: «На первый план вышли удобство и практичность». При этом всём люди стремятся создавать собственный стиль, комбинируя мировые тенденции с национальными элементами.

Социальные сети сыграли особую роль в развитии моды. Платформы позволяют дизайнерам демонстрировать свои коллекции широкой аудитории, а покупателям - быстро узнавать о новых трендах. Благодаря Интернету казахстанская мода стала более открытой и доступной для международного сообщества.

За последние годы казахстанские дизайнеры добились значительных успехов. Их работы регулярно участвуют в международных показах, конкурсах и выставках. Особенность многих коллекций заключается в сочетании современных технологий и национальных мотивов. Казахстанские дизайнеры стремятся показать миру уникальность культуры своей страны. Они используют традиционные узоры, исторические образы и элементы народного искусства, адаптируя их к современным требованиям. [6] Благодаря этому создаётся узнаваемый стиль, который привлекает внимание зарубежных экспертов и покупателей.

Развитие дизайнерской школы способствует росту интереса к отечественным брендам. Всё больше казахстанцев выбирают продукцию местных производителей,

поддерживая национальную индустрию моды. Это помогает развитию малого и среднего бизнеса, создаёт новые рабочие места и укрепляет культурный имидж страны. Одними из самых известных казахских брендов одежды являются Qazaq Republic, Kaidarova, Global Nomads и другие. [9]

Важным фактором развития индустрии являются модные показы, выставки и фестивали. Они дают возможность молодым дизайнерам представить свои работы широкой аудитории, получить профессиональную оценку и найти новых партнёров. Например, Kazakhstan Fashion Week считается одной из крупнейших модных платформ Центральной Азии и проводится с 2004 года. [4] В кастингах данного мероприятия принимают участия более 600 моделей, а в разные сезоны на показах участвуют более 15-25 дизайнеров из Казахстана и других стран региона. [7]

Подобные мероприятия привлекают внимание не только специалистов, но и обычных зрителей. Они помогают популяризировать моду как часть культуры и искусства. Кроме того, показы становятся площадкой для обсуждения современных тенденций, инновационных технологий и вопросов устойчивого развития. Модные события также способствуют развитию туризма и международного сотрудничества. Гости из других стран знакомятся с культурой Казахстана через творчество дизайнеров и получают возможность увидеть современные достижения страны.

Одним из главных мировых трендов последних лет стала экологичная мода. Всё больше производителей задумываются о влиянии индустрии на окружающую среду. Казахстан также постепенно включается в этот процесс. Многие дизайнеры начинают использовать экологически безопасные материалы, сокращать количество отходов и поддерживать принципы ответственного потребления. Покупатели всё чаще интересуются происхождением одежды и условиями её производства.

Особое значение приобретают традиционные ремёсла. Изделия ручной работы отличаются высоким качеством и долговечностью, что соответствует идеям устойчивой моды. Возрождение народных технологий помогает одновременно сохранять культурное наследие и заботиться об окружающей среде. По словам основателя Kazakhstan Fashion Week Саята Досыбаева, одной из ключевых тем последних сезонов стало уважение к природе и осознанное потребление одежды. Одновременно дизайнеры стремятся переосмыслить национальное наследие. Как отмечает Сакен Жаксыбаев, наличие традиционных элементов ещё не делает одежду по-настоящему казахстанской - современный стиль формируется через глубокое понимание культуры и её адаптацию к требованиям XXI века. [10]

Молодое поколение играет ключевую роль в развитии модной индустрии. Именно молодёжь быстрее всего воспринимает новые идеи и технологии. Сегодня многие школьники и студенты интересуются дизайном одежды, фотографией, моделингом и маркетингом в сфере моды. Современные образовательные программы, мастер-классы и творческие проекты позволяют молодым специалистам получать необходимые знания и навыки. [8] Интернет открывает доступ к международному опыту и помогает развивать собственные проекты.

В будущем казахстанская мода может стать ещё более заметной на мировой арене. Для этого важно поддерживать талантливых дизайнеров, развивать образовательные программы и сохранять связь между современными тенденциями и национальной культурой. [11]

Мода в Казахстане представляет собой уникальное сочетание исторического наследия и современных мировых тенденций. Национальные орнаменты, традиционные материалы и культурные символы продолжают вдохновлять дизайнеров, помогая создавать узнаваемый стиль. Одновременно влияние глобальной модной индустрии способствует внедрению новых технологий и развитию творческих направлений. Сегодня казахстанская мода является не только частью экономики, но и важным инструментом сохранения культурной идентичности. Благодаря работе дизайнеров, развитию модных мероприятий и интересу молодёжи страна формирует собственный образ в мировом культурном пространстве. Именно поэтому мода в Казахстане остаётся одним из наиболее интересных примеров того, как традиции и современность могут успешно сосуществовать и дополнять друг друга.

Список использованной литературы:

1. История Казахстана: учебник для общеобразовательных школ. – Астана: Министерство просвещения РК, 2023.
2. История казахского национального костюма / под ред. К. Байпакова. – Алматы: Қазақ университеті, 2020.
3. Ергалиева Р. Традиционная культура казахского народа. – Алматы: Аруна, 2019.
4. Материалы официального сайта Kazakhstan Fashion Week.
5. Материалы официального сайта Almaty Fashion Week.
6. Электронный журнал Marie Claire Kazakhstan.
7. Информационно-аналитический портал The Steppe.
8. Деловое издание Inbusiness.kz.
9. Официальные сайты казахстанских брендов одежды Qazaq Republic, ARUNAZ, Kaidarova, Global Nomads.
10. Материалы открытых интервью казахстанских дизайнеров и организаторов модных мероприятий за 2024–2026 годы.
11. Интернет-ресурсы по истории казахского национального костюма и современным тенденциям моды в Казахстане.

UDC 697.1:502.131

Alikhan Bulekbayev

Nazarbayev Intellectual School of Physics and Mathematics
(Uralsk, Kazakhstan)
Scientific supervisor: Huy D. Le,
PhD student
Columbia University

TO WHAT EXTENT DO GREEN ROOFS MITIGATE ROOFTOP SURFACE AND AMBIENT AIR TEMPERATURES ACROSS SEASONAL EXTREMES IN KAZAKHSTAN? A CASE STUDY OF ALMATY

Abstract: The aim of the study is to analyze the effect of green roofs on surface and air temperatures in the continental climate of Kazakhstan using the city of Almaty as an example. Green roofs are roofs of buildings covered with vegetation that cools the roofs by absorbing heat and releasing it into the environment. Based on satellite data analysis, a stable cooling capacity of green roofs was found in early summer, with a cooling effect of 3–4 °C compared to conventional roofs. In contrast, the cooling effect was less pronounced on extreme heat days, when in some cases the green roof recorded higher temperatures. In winter, on the contrary, green roofs remained cooler than conventional roofs both during the day and at night, which could lead to additional heat loss. The results of this study show that green roofs can contribute to climate mitigation in Almaty and possibly other cities in Kazakhstan, provided that their design features are adapted to extreme weather conditions. At the same time, the cooling effect in winter indicates their ability to increase heating demand, which highlights the dual nature of their impact and the need to take climate conditions into account when designing. The article outlines areas for further research, including physical measurements, modeling, and adaptation and optimization of designs taking into account local climate factors.

Keywords: green roofs; urban heat island; remote sensing; continental climate; Kazakhstan; sustainable urban infrastructure

1. Introduction

Climate change has intensified the frequency and severity of extreme weather events, especially heatwaves, which are becoming increasingly life-threatening. One of the most pertinent effects is now considered the urban heat island (UHI) effect that keeps heat in densely-built up urban areas with respect to its rural counterpart , thereby intensifying urban heatwaves, increasing cooling energy demand, and exposing vulnerable populations to health risks. To mitigate UHI effects, green roofs — layer of vegetation installed on the top of buildings, been increasingly implemented in cities worldwide. Past studies suggest that green roofs in temperate and tropical climate conditions can lower surface temperatures by 20–30 °C and reduce air temperature by 1–2 °C ((Jamei, E., Rajagopalan, P., Seyedmahmoudian, M., & Jamei, Y. (2021). The impact of green roofs on urban heat island: A meta-analysis. *Building and Environment, 196,* 107790.)). The potential of green roofs in maintaining low temperatures has not yet been fully evaluated for the sharply continental climate of Central Asia. In summer, temperatures can rise above 35 °C, while in winter they can drop to -20 °C, in a place such as

Almaty, Kazakhstan. Although simulations suggested such possibilities ((Kulumkanov, A., Bekmukhambetov, Y., & Zhaksylykova, K. (2023). Green roof energy performance in different climate zones: a simulation study)), the actual satellite or in-situ observations are yet to ascertain such benefits. This gap in the study is important because it is still unclear how green roofs will perform during extreme seasons in Kazakhstan.

To address this gap, the present study seeks to determine, in an empirical manner, the influences that green roofs have on rooftop surface temperature (Ts) and estimated ambient air temperature (Ta) through extremes of summer and winter Almaty, Kazakhstan. We selected Almaty as the study site due to its status as the largest urban center in Kazakhstan and the only city in the country where green roofs have been implemented. The main objectives are:

- To analyze existing literature on the effectiveness of green roofs.
- To model rooftop surface and ambient air temperature of green and normal roofs with the help of Landsat 8.
- To conduct phase quantitative analyses and draw inferences.

This study concentrated on six buildings in Almaty, with three being green-roofed and three normal-roofed buildings, and was monitored through Landsat 8 satellite imagery from June 2023 to February 2024. Both the daytime and nighttime thermal data were analyzed, and air temperature was estimated through an empirically derived regression model. While the analysis yielded interesting results on the thermal characteristics of roofs, it remains limited due to the lack of in-situ measurements, the possible variation in roof insulation, and the relatively small number of buildings used as samples. These results were visualized through temperature anomalies and discussed in relation to seasonal performance.

2. Methods

2.1. Literature Review

To provide background to the main study, a focused review of literature was undertaken, selecting peer-reviewed articles published between 2021 and 2025. Searches were carried out on Google Scholar, ResearchGate and ScienceDirect, combining such keywords as "green roofs," "urban heat island," "continental climate," "thermal performance," and "Kazakhstan." Studies were considered relevant only if they dealt with some aspect of green roofing in hot, dry, semi-arid, or continental climates with emphasis on the effect of green roofs on rooftop or urban temperature. The review sought to draw data on methodologies used, key conclusions of studies, climatic settings, and limitations, focusing on whether or not the study contained empirical analysis at the rooftop level.

Jamei et al. (2021) performed an overview of global studies of the green roofs' effectiveness, and the results of 89 peer-reviewed works from 2000 to 2020 were synthesized. The findings evidenced temperature reduction of up to 30°C on the rooftops and 3°C across ambient air temperature, with exceptional results under hot and dry climate conditions ((Jamei et al. (2021) performed an overview of global studies of the green roofs' effectiveness, and the results of 89 peer-reviewed works from 2000 to 2020 were synthesized. The findings evidenced temperature reduction of up to 30°C on the rooftops and 3°C across ambient air temperature, with exceptional results under hot and dry climate conditions.)). The author emphasizes that green roofs are a very effective tool for dealing with overheating and notes that they are less effective in continental climates. However, they emphasized that while green roofs are generally effective in reducing urban overheating, their performance tends to decline in continental climates, which remain underexamined in the existing literature.

Sahnoune et al. (2021) focused on the city of Constantine in Algeria, which enjoys a semi-arid climate, as do some regions of Kazakhstan ((Sahnoune, F., Bouzrara, M., & Belarbi, R. (2021). Assessment of green roof cooling potential in hot and arid climates: Case study of Algeria. *Journal of Building Engineering, 43,* 102918.)). With the help of Landsat 8 satellite imagery and GIS (Geographic Information System) modeling, the researchers found that green roofs and other green infrastructure can develop a strong cooling effect that reduces surface temperatures by about 4°C and air temperatures by 1.2°C during extreme heat. Research validates the approach of satellite monitoring to determine how effective green solutions are at cooling dry urban environments. However, unlike Jamei's global synthesis, Sahnoune's study centered on one urban setting, aggregating cooling effects across city zones without distinguishing green roofs from other greenery ((Jamei, E., Rajagopalan, P., Seyedmahmoudian, M., & Jamei, Y. (2021). The impact of green roofs on urban heat island: A meta-analysis. *Building and Environment, 196,* 107790.))' ((Sahnoune, F., Bouzrara, M., & Belarbi, R. (2021). Assessment of green roof cooling potential in hot and arid climates: Case study of Algeria. *Journal of Building Engineering, 43,* 102918.)). This prevents any direct comparison and puts in perspective the geographical dimension; for instance, Jamei's review underlined the lack of data at the continental level, while algerian's regional case-study seemingly confirms potential, yet without the ability to drill down to the rooftop level ((Jamei, E., Rajagopalan, P., Seyedmahmoudian, M., & Jamei, Y. (2021). The impact of green roofs on urban heat island: A meta-analysis. *Building and Environment, 196,* 107790.)).

Regional research concentrated on Kazakhstan-specific modeling for simulating the impact of green roofs on energy performance in cities like Astana and Almaty ((Kulumkanov, A., Bekmukhambetov, Y., & Zhaksylykova, K. (2023). Green roof energy performance in different climate zones: a simulation study).) Their study focused on using building energy simulation tools to quantify the potential of green roof application in reducing cooling loads and enhancing thermal comfort under arid continental conditions. The result confirmed that this type of green roof could reduce energy consumption, mostly during the summer. Jamei et al. showed the continental gap; Sahnoune et al. tried remote sensing but lacked rooftop-specific targeting ((Jamei, E., Rajagopalan, P., Seyedmahmoudian, M., & Jamei, Y. (2021). The impact of green roofs on urban heat island: A meta-analysis. *Building and Environment, 196,* 107790.))' ((Sahnoune, F., Bouzrara, M., & Belarbi, R. (2021). Assessment of green roof cooling potential in hot and arid climates: Case study of Algeria. *Journal of Building Engineering, 43,* 102918.)). By contrast, regional research showed theoretical promise in the same region but cannot ascertain real-world temperature effects. Hence, while all three studies seem to converge on the potential of green roofs, they differ in their method, level of spatial resolution, and kind of evidence used — global meta-analysis, city-level remote sensing, and simulated modeling, respectively.

Taken collectively, these studies reveal a critical scientific gap: no empirical study has used satellite observations to directly compare surface and air temperature performance of green versus conventional roofs in Kazakhstan's continental climate. This is particularly important given the unique seasonal extremes of Almaty. A review of the literature confirmed the high relevance of this topic and demonstrated that our research can fill existing gap.

2.2. Remote Sensing Temperature Data Analysis

For analyzing the efficiency of green roofs, we decided to use satellite data analysis, because it allows for a large-scale and consistent comparison of land surface temperatures across

different rooftop types. We employed Google Earth Engine (GEE) as a primary platform. The use of Google Earth Engine (GEE) was paramount in our study for processing and visualizing Landsat 8 imagery. Landsat 8 carries thermal infrared bands (especially Band 10) used for LST (Land Surface Temperature) calculation. With these, surface temperatures of green roofs and conventional roofs are compared on the same dates while accounting for having equal environmental conditions. The city of Almaty was selected as the subject of the study, as it has the largest number of green roofs and a growing building density. The analysis covered six representative buildings in the central part of the city—three with green roofs and three with conventional roofs. For the analysis, cloud-free satellite images were selected for the 15th day of each characteristic month (June, July, August, December, January, and February 2023–2024) to reflect seasonal differences under similar solar conditions. In addition, measurements were taken for both daytime and nighttime passes, allowing for a more comprehensive understanding of temperature dynamics. Ambient air temperature had to be estimated, as surface temperature is not equal to air temperature, especially in the urban environment. Landsat 8 gives very accurate data of LST, which is actually a measure of heat on rooftops or relatively flat ground surfaces. It does not signify air temperature felt by humans or the air temperature around them.. Calculating air temperature was important because it affects people's health and comfort. To do this, an empirical formula was used, taken from a study by Janghoo L. ((Janghoo, L. (2025). Estimating Near-Surface Air Temperature from Satellite-Derived Land Surface Temperature Using Temporal Deep Learning: A Comparative Analysis)), which calculates air temperature based on the land surface temperature (LST)

$$Ta = a \cdot LST + b$$

Where:

Ta - is the estimated air temperature (°C)
LST- is the land surface temperature (°C) derived from satellite data
a and b are regression coefficients determined empirically for specific geographic and climatic conditions

In this study, the parameters a=0.89 and b=-1.98 were adopted by the results of regression analysis for moderately continental urban conditions ((Janghoo, L. (2025). Estimating Near-Surface Air Temperature from Satellite-Derived Land Surface Temperature Using Temporal Deep Learning: A Comparative Analysis)). These coefficients were selected since the climate of Almaty—hot summers, cold winters, and another moderate period of precipitation—is very much similar to those of temperate continental cities used in Lee's regression model. According to the Köppen–Geiger classification ((Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen–Geiger climate classification. *Hydrology and Earth System Sciences, 11(5),* 1633–1644.)), Almaty possesses a cold continental climate (Dfa/Dfb), so the formula can be considered reasonably applicable. This allows for a fairly accurate estimate of the air temperature on roofs and its seasonality, especially when there is no data on temperature obtained from on-site measurements. As a result, we obtained an empirical formula for calculating air temperature from the earth's surface temperature, taking into account all the coefficients:

$$Ta = 0.89 * LST - 1.98$$

After constructing daily and nightly temperature series for each date, we immediately calculated the temperature anomaly defined as the difference between the temperature of normal and green

roofs. The calculation was performed separately for surface temperature and air temperature, during the day and at night. The cooling anomaly (ΔT) formula used in this study was adapted, where the temperature difference between conventional and green roofs was quantitatively assessed. The anomaly is defined as:

$$\Delta T = \frac{1}{n_o}\sum_{j=1}^{n_o} T_{o,j} - \frac{1}{n_g}\sum_{i=1}^{n_g} T_{g,i}$$

Where:

ΔT – anomaly (°C), i.e. the difference between the average temperature of ordinary roofs and green roofs.

$T_{o,j}$ – temperature on j-th ordinary roof

$T_{g,i}$ – temperature on i-th green roof

n_g – number of ordinary roofs

n_o – number of green roofs

ΔT shows the difference between the temperature of a green roof and a conventional roof. If ΔT is negative, it means that the green roof is cooler than a conventional roof — that is, it effectively reduces the temperature, as intended. If ΔT is positive, on the contrary, the green roof is warmer than a conventional roof, which may indicate that it is not currently functioning as a cooling mechanism. Thus, the ΔT indicates whether green roofs truly help combat overheating in buildings under various weather conditions. After calculating all the data, we compiled the results into tables and constructed graphs that reflect seasonal differences and the comparative effectiveness of green and conventional roofs.

Table 1. Surface Temperatures of Green and Conventional Roofs in Almaty Across Seasonal Extremes (Landsat 8, 2023-2024)

June 15				
Day			Night	
Test	Roof surface temperature (°C)	Air temperature (°C)	Land surface temperature (°C)	Air temperature (°C)
Green 1	33.26	28.6	25.98	21.10
Green 2	35.13	28.2	27.10	22.21
Green 3	34.16	28.9	26.43	21.57
Normal 1	37.08	32.0	30.05	24.74
Normal 2	38.52	31.4	31.27	25.86

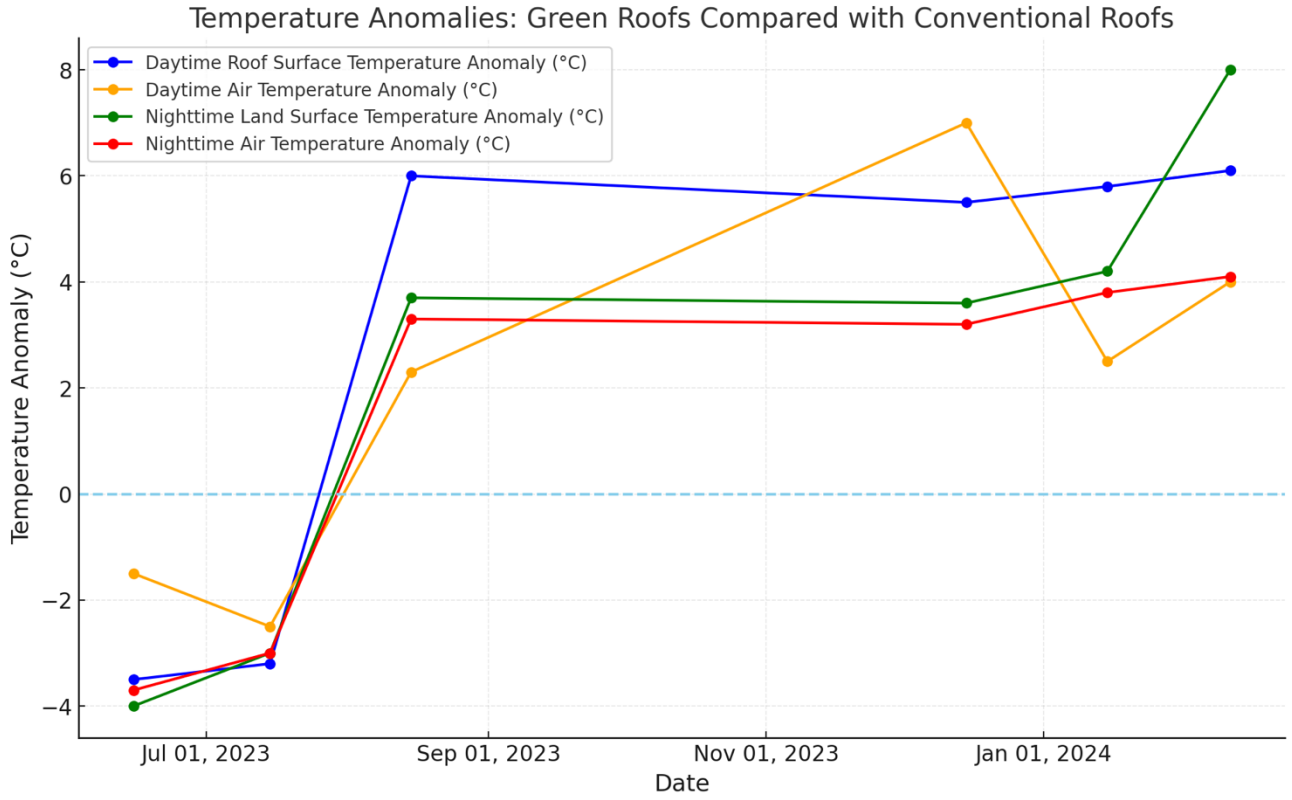
Normal 3	37.60	30.6	30.95	25.58
	Anomaly(°C)= -3.55	Anomaly(°C)= -2.76	Anomaly(°C)= -4.26	Anomaly(°C)= -3.76
July 15				
Green 1	42.06	36.1	36.5	31.6
Green 2	43.84	35.9	38.3	31.0
Green 3	41.65	36.3	36.0	31.2
Normal 1	44.59	39.2	39.2	34.3
Normal 2	46.10	39.8	40.5	34.9
Normal 3	46.18	39.9	40.6	35.0
	Anomaly(°C)= -3.10	Anomaly(°C)= -3.53	Anomaly(°C)= -3.17	Anomaly(°C)= -3.46
August 15				
Green 1	39.06	42.0	36.5	38.5
Green 2	40.84	43.5	38.2	39.8
Green 3	38.65	41.2	35.9	39.8
Normal 1	44.59	39.5	39.8	35.5
Normal 2	46.10	40.2	41.0	36.2

Normal 3	46.18	40.0	41.2	36.0
	Anomaly(°C)= +6.10	Anomaly(°C)= +2.33	Anomaly(°C)= +3.80	Anomaly(°C)= +3.47
December 15				
Green 1	-5.00	-13.10	-6.2	-7.5
Green 2	-4.40	-12.4	-5.9	-7.2
Green 3	-6.10	-14.1	-7.4	-8.6
Normal 1	-0.50	-6.5	-3.5	-5.0
Normal 2	0.30	-5.7	-2.8	-4.3
Normal 3	0.80	-5.2	-2.1	-3.8
	Anomaly(°C)= +5.37	Anomaly(°C)= +7.4	Anomaly(°C)= +3.7	Anomaly(°C)=+3.4
January 15				
Green 1	-7.00	-2.5	-8.4	-9.4
Green 2	-6.40	-2.0	-7.7	-8.8
Green 3	-8.10	-3.0	-9.9	-10.8
Normal 1	-1.80	-0.5	-4.5	-6.0
Normal 2	-1.00	+0.2	-3.8	-5.2

Normal 3	-0.30	+0.5	-3.0	-4.7
	Anomaly(°C)=+6.14	Anomaly(°C)= +2.57	Anomaly(°C)= +4.90	Anomaly(°C)= +4.37
February 11				
Green 1	-8.00	-3.80	-9.6	-10.5
Green 2	-7.20	-3.10	-8.9	-9.9
Green 3	-9.00	-4.5	-10.4	-11.2
Normal 1	-2.50	+0.2	-5.5	-6.9
Normal 2	-2.00	+0.5	4.8	-6.2
Normal 3	-1.30	1.0	-4.1	-5.7
	Anomaly(°C)= +6.14	Anomaly(°C)= +4.37	Anomaly(°C)= +8.03	Anomaly(°C)= +4.26

Table 2. Locations and Roof Characteristics of Surveyed Buildings in Almaty

Experiment	Name of building	Type of roof	Latitude	Longitude
Green 1	Park View Office Tower	Green	43.250205° N	76.945112° E
Green 2	DoubleTree Hilton	Green	43.244810° N	76.939510° E
Green 3	Hilton Roof Garden	Green	43.244810° N	76.939510° E
Normal 1	Nurlu Tau	Normal	43.243259° N	76.950439° E
Normal 2	Bukhar Zhyrau Towers	Normal	43.243360° N	76.926356° E
Normal 3	Ramada Almaty	Normal	43.249430° N	76.937671° E



Pic. 1. Temperature Anomalies

3. Discussion

Early summer observations also verified the cooling impact of green roofs. The coverings of green roofs were 3.55 °C cooler than non-green roofs during daytime on June 15th, and the air above them was 2.76 °C cooler. During nighttime, this gap widened to 4.26 °C (Ts) and 3.76 °C (Ta). The same trend was seen on the 15th of July, where the green roofs reduced in temperature by 3.10 °C (Ts) and 3.50 °C (Ta) during the day, and 3.17 °C and 3.46 °C during the night. On August 15, however, during the hottest heat wave, a different trend was seen. Green roof surfaces increased by 6.10 °C relative to the normal roofs, and the air above them by 2.33 °C. At night, the excess heat was mitigated but remained at 3.80 °C (Ts) and 3.47 °C (Ta). Thus, while green roofs maintained a uniform 3–4 °C cooling during early summer, their performance declined and even became negative when exposed to extreme heat. Under winter conditions, green roofs were always cooler than the reference roofs. On December 15, traditional roofs outperformed green roofs by 5.37 °C (Ts) and 7.40 °C (Ta) during the day and by 3.70 °C and 3.40 °C at night, respectively. On January 15, the differences between them increased to 6.14 °C (Ts) and 2.57 °C (Ta) during the day and 4.90 °C and 4.37 °C at night. The largest contrast between green roofs and conventional roofs was on February 15: green roofs were 8.03 °C cooler than conventional roofs at night and 6.14 °C during the day, while the air temperature was 4.26 °C and 4.37 °C higher, respectively. Green roofs were cooler both day and night during the entire winter.

The research is the first to experimentally examine the thermal performance of green roofs in the harsh continental climate of Kazakhstan, and more significantly in the context of freeze and heatwave problems in Almaty. Jamei et al. (2021) and Sahnoune et al. (2021) also conducted studies on green roofs in rainy or tropical climates but listed in their data were not those under the climatic conditions of Kazakhstan. Kazakhstani research had suggested a model for green roof performance in Kazakhstan but had theoretical findings that were not empirically verified

with real data ((Jamei, E., Rajagopalan, P., Seyedmahmoudian, M., & Jamei, Y. (2021). The impact of green roofs on urban heat island: A meta-analysis. *Building and Environment, 196,* 107790))' ((Sahnoune, F., Bouzrara, M., & Belarbi, R. (2021). Assessment of green roof cooling potential in hot and arid climates: Case study of Algeria. *Journal of Building Engineering, 43,* 102918.)). This research closes the gap by utilizing satellite imagery to test the actual performance of green roofs with excessive heat conditions.

Our findings in June and July concurred with Jamei et al. (2021) meta-analysis in which green roofs lowered surface and air temperature by 3 °C ((Jamei, E., Rajagopalan, P., Seyedmahmoudian, M., & Jamei, Y. (2021). The impact of green roofs on urban heat island: A meta-analysis. *Building and Environment, 196,* 107790)). Cooling is a result of mechanisms like evaporative cooling, large leaf albedo, and low thermal inertia of the vegetation layer. In the case of extreme heat in August, however, the cooling was reversed. This disparity confirms the work of Sahnoune et al. (2021), who also recorded such rises in temperature on green roofs because of substrate evaporation ((Sahnoune, F., Bouzrara, M., & Belarbi, R. (2021). Assessment of green roof cooling potential in hot and arid climates: Case study of Algeria. *Journal of Building Engineering, 43,* 102918.)). At temperatures higher than 35–40 °C and in the absence of any moisture, evaporative cooling ceases and the stored water in the soil is absorbed and re-radiated, leading to an increase in temperature. Our investigation indicates that even drought-tolerant plants such as sedums, used in both Almaty and Algeria, cannot cool without water. Green roofs under winter were cooler than control roofs at all times, including nighttime, with temperature differences of up to 8°C. This suggests that the green roof cooling effect endures during the colder periods, and this is a potential benefit in the sense of lower rooftop temperatures than standard roofs, especially during highly seasonal locations such as Almaty.

Findings of this study dictate some practical implications. To preserve yearly cooling effects, green roofs in continental climates must include irrigation systems, drought vegetation, and more substantial substrates. Particular attention should also be paid to winter, when green roofs lose more heat compared to conventional roofs, which could require extra insulation. Urban planning should account not only for the green roof cooling effect in the summer but also the winter performance of green roofs. There are, however, some constraints of the study. The data were obtained from satellite observations on a monthly basis, thus the temporal resolution of the temperature records was limited. Additionally, the satellite data only gives surface temperature, from which air temperature is being estimated using a regression model, and this can be inaccurate. Secondly, the research was conducted from a limited sample of six buildings that might not reflect the variety of green roof performance in various urban spaces in Kazakhstan.

Future research should focus on incorporating ground-truth measurements of temperature, humidity, and soil moisture to further validate the findings of this study. Such research could provide insights into the specific thresholds at which green roofs cease to be effective and begin to act as heat sources. Additionally, further studies could explore different vegetation types, soil depths, and irrigation strategies to optimize the design of green roofs for extreme climate conditions.

4. Acknowledgments

I would like to thank Huy D. Le for his guidance and mentorship throughout the development of this research paper.

References:

1. Jamei, E., Rajagopalan, P., Seyedmahmoudian, M., & Jamei, Y. (2021). The impact of green roofs on urban heat island: A meta-analysis. Building and Environment, 196, 107790.
2. Janghoo, L. (2025). Estimating Near-Surface Air Temperature from Satellite-Derived Land Surface Temperature Using Temporal Deep Learning: A Comparative Analysis.
3. Kulumkanov, A., Bekmukhambetov, Y., & Zhaksylykova, K. (2023). Green roof energy performance in different climate zones: a simulation study.
4. Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen–Geiger climate classification. Hydrology and Earth System Sciences, 11(5), 1633–1644.
5. Sahnoune, F., Bouzrara, M., & Belarbi, R. (2021). Assessment of green roof cooling potential in hot and arid climates: Case study of Algeria. Journal of Building Engineering, 43, 102918.

UDC 612.821:613.71

Takhmina-Sultanbekova

Grade 10 student
physics and mathematics track
Scientific supervisor: Aksenova Inna Valerievna
Nazarbayev Intellectual School (NIS)
(Taraz, Kazakhstan)

PHYSICAL ACTIVITY IS ASSOCIATED WITH PROCESSING SPEED AND VERBAL FLUENCY, BUT NOT VERBAL MEMORY, IN OLDER ADULTS: A DOMAIN-SPECIFIC ANALYSIS OF NHANES 2011–2014

Abstract: Physical activity is widely linked to better cognition in older adults, but most population studies use a single global cognitive score, or analyze several cognitive tests separately without checking whether the strength of the association differs from one cognitive domain to another. Using the U.S. National Health and Nutrition Examination Survey (NHANES) 2011–2014, this study examined whether meeting recreational physical-activity guidelines is associated with three distinct cognitive domains in adults aged 60 and older: processing speed (Digit Symbol Substitution Test), semantic fluency (Animal Fluency), and verbal memory (CERAD Word Learning), and then tested directly whether the associations differ across these domains. In survey-weighted, fully adjusted models (n = 2,568), meeting the recreational physical-activity guideline was associated with higher semantic fluency (standardized $\beta = +0.28$) and faster processing speed ($\beta = +0.15$), but not with verbal memory ($\beta = +0.02$). A formal cross-domain test confirmed that the association differs significantly across the three domains ($F = 35.6$, $p < 0.001$). The pattern held with a continuous activity measure and with an alternative memory test, and did not differ between men and women. The cross-sectional cognitive correlates of physical activity in older adults are therefore domain-selective: they concentrate in fluency and processing speed rather than in verbal memory. Because the design is cross-sectional, the findings describe association rather than cause.

Keywords: physical activity, cognitive function, older adults, NHANES, processing speed, verbal fluency, memory

Introduction. Populations around the world are aging, and protecting cognitive function in later life has become a major public-health goal. Among the factors people can change, physical activity is one of the most studied in relation to cognitive health in older adults. Meta-analyses of exercise programs report small-to-moderate benefits for cognition in adults over 50 [1], and laboratory and imaging studies connect physical activity to brain structure and function, including hippocampal volume and the control processes of the frontal lobes [2, 3, 4, 5].

Cognition, however, is not a single ability. Neuropsychology separates it into domains such as processing speed, executive function, verbal fluency, and episodic memory, and these domains rely on partly different brain systems. This raises a question that goes beyond whether physical activity relates to cognition at all: which cognitive domains does it relate to most strongly? If the association is concentrated in particular domains, that pattern matters both for

understanding how physical activity might act on the brain and for choosing which cognitive outcomes to measure in future studies.

NHANES is well suited to this question. For participants aged 60 and older it administered three validated tests covering distinct domains (the CERAD Word Learning test for verbal memory, the Animal Fluency test for semantic fluency, and the Digit Symbol Substitution Test for processing speed), together with a standardized physical-activity questionnaire and detailed health and demographic measures.

Earlier work on these exact NHANES cycles examined occupational, transport, and recreational physical activity in relation to each cognitive test [6]. That study reported associations test by test, finding that recreational and total activity related to fluency and processing speed, with weaker or null associations for memory. But it analyzed each cognitive outcome in a separate model and never tested whether the strength of the association differs across cognitive domains. The gradient was visible in its tables, yet it was never put to a statistical test. The present study closes that gap: the three cognitive outcomes are placed on a common scale and a single pooled model tests and ranks the association of physical activity across domains.

The objective was to examine whether meeting recreational physical-activity guidelines is associated with processing speed, semantic fluency, and verbal memory in U.S. adults aged 60 and older, and whether these associations differ across domains. We hypothesized a gradient in which physical activity is more strongly associated with processing speed and fluency than with verbal memory.

Materials and Methods. Data source and population. The study used public data from NHANES, a nationally representative, continuously fielded survey of the civilian U.S. population conducted by the National Center for Health Statistics [7]. NHANES uses a complex, multistage probability sample. The two consecutive cycles 2011–2012 and 2013–2014 were pooled because these are the cycles in which the adult cognitive battery was administered. That battery was given to participants aged 60 and older, so the sample was restricted to that age group. After excluding participants with missing data on any analysis variable, the analytic sample was 2,568 adults.

Cognitive outcomes. Three tests were used, each representing a different domain. Processing speed was measured with the Digit Symbol Substitution Test, in which participants match symbols to numbers against a time limit; higher scores indicate faster processing speed and sustained attention [8]. Semantic fluency was measured with the Animal Fluency test, the number of unique animals named in 60 seconds. Verbal memory was measured with CERAD Word Learning, the immediate-recall total across three learning trials of a 10-word list [9]; delayed recall was used in a sensitivity analysis. To compare associations across domains on a common scale, each cognitive score was converted to a z-score (mean 0, standard deviation 1) within the analytic sample.

Physical activity. Physical activity was measured with the Global Physical Activity Questionnaire, which records the frequency and duration of vigorous and moderate activity across work, transport, and recreational settings [10]. Because occupational and transport activity show weak or null associations with cognition in this population [6], the analysis focused on recreational activity, expressed in MET-minutes per week. The primary exposure was a binary indicator of meeting the recreational activity guideline (at least 600 MET-minutes

per week, consistent with WHO guidance [11]). A continuous measure (log MET-minutes per week) was used as a sensitivity analysis.

Covariates. Fully adjusted models controlled for age, sex, education, family income-to-poverty ratio, race and ethnicity, survey cycle, depressive symptoms (PHQ-9), diabetes, hypertension, body mass index, and smoking status. These variables are associated with both physical activity and cognition and could otherwise confound the relationship.

Statistical analysis. All analyses incorporated the NHANES sampling design through examination weights (halved across the two pooled cycles) and design-based standard errors clustered on primary sampling units within strata. A separate survey-weighted linear regression was first fit for each standardized cognitive domain on the physical-activity indicator plus covariates, giving three standardized coefficients. To test whether the association differs across domains, the central question of this study, the data were then reshaped into a long person-by-domain structure and a single pooled model was fit containing a physical-activity by domain interaction, with variance clustered at the design level. A joint Wald test of the interaction terms tested the null hypothesis that the association is equal across the three domains. The analysis plan, including this cross-domain test, was written before the results were examined. Secondary analyses tested whether the association differed between men and women; sensitivity analyses used the continuous activity measure and CERAD delayed recall as an alternative memory outcome. A two-sided $p < 0.05$ was treated as significant.

Ethics. NHANES is conducted under National Center for Health Statistics Research Ethics Review Board approval, and all participants provided informed consent. This secondary analysis used only de-identified public data and required no additional ethical approval.

Results. Sample characteristics. Among the 2,568 older adults, the mean age was about 69 years and 53% were women. About 27% met the recreational physical-activity guideline. Those who met it were, on average, slightly younger and had higher income, lower body mass index, and a lower prevalence of diabetes than those who did not (Table 1), which is why statistical adjustment matters here. Their unadjusted cognitive scores were also higher on all three tests.

Table 1 – Weighted characteristics by physical-activity guideline status.

Characteristic	Below guideline (n = 1,865)	Meets guideline (n = 703)
Age, years	69.4	68.5
Female, %	56	49
Income-to-poverty ratio	2.94	3.49
Diabetes, %	22	13
Body mass index	30.0	27.4
DSST (processing speed), points	50.0	57.1
Animal Fluency, words	17.4	20.1
CERAD immediate recall, points	19.5	20.2

Domain-specific associations. In fully adjusted, survey-weighted models, meeting the recreational physical-activity guideline was associated with higher cognitive scores in two of the three domains (Figure 1): semantic fluency ($\beta = +0.28$ SD, 95% CI 0.17 to 0.39; $p < 0.001$) and processing speed ($\beta = +0.15$ SD, 95% CI 0.05 to 0.25; $p = 0.004$), but not verbal memory

(β = +0.02 SD, 95% CI −0.06 to 0.11; p = 0.60). The confidence interval for memory rules out effects larger than about 0.11 SD, so the null is reasonably precise rather than simply underpowered.

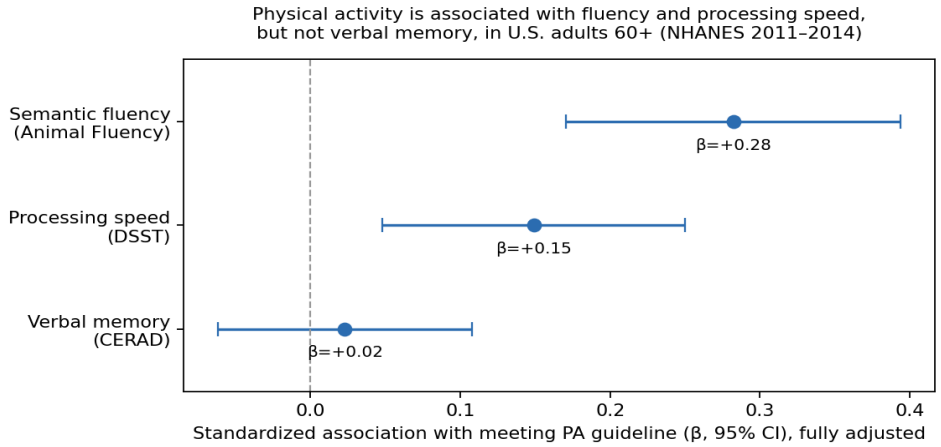


Figure 1 – Standardized association between meeting the recreational physical-activity guideline and each cognitive domain (fully adjusted, with 95% confidence intervals).

Formal cross-domain test. The pooled model with a physical-activity by domain interaction rejected the hypothesis that the association is equal across domains (joint Wald test F = 35.6, p < 0.001). Relative to verbal memory, the association was significantly larger for both semantic fluency and processing speed. This is the central result: in older adults, the link between physical activity and cognition is domain-selective, not uniform.

Secondary and sensitivity analyses. The continuous physical-activity measure produced the same ordering (fluency, then processing speed, then memory, with memory non-significant). Using CERAD delayed recall as an alternative memory outcome, the association with physical activity remained non-significant (β = +0.08, p = 0.13), so the memory null is not specific to one memory measure. The association did not differ significantly between men and women in any domain (all interaction p ≥ 0.18), so the gradient looks similar in both sexes.

Discussion. In a large, nationally representative sample of older U.S. adults, meeting recreational physical-activity guidelines was associated with better semantic fluency and faster processing speed, but showed essentially no association with verbal memory, and a formal test confirmed that these domain differences are real. The contribution here is not that a physical-activity and cognition association exists, which is well established, but that the association is domain-selective and can be shown to be so with an explicit cross-domain test.

This finding builds on, and differs from, earlier work on the same NHANES cycles. Liang and colleagues [6] reported physical-activity associations test by test and saw a similar pattern, but did not test whether the domains differed, so the gradient was visible without being established. By standardizing the outcomes and testing the interaction directly, this study moves the claim from "physical activity relates to several cognitive tests to varying degrees" to "physical activity relates to cognition differently across domains, and significantly so."

The pattern fits a broader literature in which the cognitive correlates of physical activity and fitness are strongest for executive and speed-related processes [3, 4, 5]. Within that pattern, the association here was numerically strongest for semantic fluency rather than for the processing-speed measure, a slight departure from the common expectation that speed-based tests are the most sensitive to physical activity; both domains, however, separated clearly from

verbal memory. Semantic fluency draws on both executive control and processing speed, and the Digit Symbol Substitution Test is a classic measure of speed and attention. Both may be more sensitive than a word-list recall task to the cardiovascular and frontal-lobe pathways through which physical activity is thought to act. This mechanistic reading is offered as context rather than as something the present data can prove.

The practical reading should stay modest. The results suggest that, when researchers study how physical activity relates to thinking in older adults, fluency and processing-speed measures may be more informative than a single memory test, and that public-health messages linking activity to "brain health" are most defensible for speed and executive-type abilities. The data do not show that becoming more active will improve any specific cognitive domain, and they should not be read that way.

This study has several limitations. First, the design is cross-sectional, so the results describe association, not cause, and reverse causation is plausible: adults with sharper processing speed and executive function may be more likely to take up recreational activity. Second, physical activity was self-reported and is subject to recall and social-desirability bias, which can blur associations. Third, the findings apply to U.S. adults aged 60 and older and may not transfer to other ages or countries. Fourth, despite extensive adjustment, residual and unmeasured confounding, for example diet, social engagement, or early undiagnosed disease, cannot be ruled out. Finally, each domain was measured with a single test; convergent measures would strengthen the domain inferences.

Conclusion. Among older U.S. adults, the cross-sectional association between recreational physical activity and cognition is domain-selective: it is present for semantic fluency and processing speed, absent for verbal memory, and the difference across domains is statistically significant. Longitudinal and interventional studies can test whether this selectivity reflects a causal, domain-specific effect of physical activity, and whether choosing fluency and processing-speed outcomes makes physical-activity studies in older adults more sensitive.

References:

1. Northey J.M., Cherbuin N., Pumpa K.L., Smee D.J., Rattray B. Exercise interventions for cognitive function in adults older than 50: a systematic review with meta-analysis // British Journal of Sports Medicine. – 2018. – Vol. 52, No. 3. – P. 154–160. – DOI: 10.1136/bjsports-2016-096587.

2. Hillman C.H., Erickson K.I., Kramer A.F. Be smart, exercise your heart: exercise effects on brain and cognition // Nature Reviews Neuroscience. – 2008. – Vol. 9, No. 1. – P. 58–65. – DOI: 10.1038/nrn2298.

3. Colcombe S., Kramer A.F. Fitness effects on the cognitive function of older adults: a meta-analytic study // Psychological Science. – 2003. – Vol. 14, No. 2. – P. 125–130. – DOI: 10.1111/1467-9280.t01-1-01430.

4. Smith P.J., Blumenthal J.A., Hoffman B.M. et al. Aerobic exercise and neurocognitive performance: a meta-analytic review of randomized controlled trials // Psychosomatic Medicine. – 2010. – Vol. 72, No. 3. – P. 239–252. – DOI: 10.1097/PSY.0b013e3181d14633.

5. Erickson K.I., Voss M.W., Prakash R.S. et al. Exercise training increases size of hippocampus and improves memory // Proceedings of the National Academy of Sciences. – 2011. – Vol. 108, No. 7. – P. 3017–3022. – DOI: 10.1073/pnas.1015950108.

6. Liang J. et al. Relationship between domain-specific physical activity and cognitive function in older adults – findings from NHANES 2011–2014 // Frontiers in Public Health. – 2024. – Vol. 12. – Art. 1390511. – DOI: 10.3389/fpubh.2024.1390511.

7. Centers for Disease Control and Prevention, National Center for Health Statistics. National Health and Nutrition Examination Survey (NHANES) 2011–2012, 2013–2014: public data files [Electronic resource]. URL: https://www.cdc.gov/nchs/nhanes/ (accessed: 25.06.2026).

8. Jaeger J. Digit Symbol Substitution Test: the case for sensitivity over specificity in neuropsychological testing // Journal of Clinical Psychopharmacology. – 2018. – Vol. 38, No. 5. – P. 513–519. – DOI: 10.1097/JCP.0000000000000941.

9. Morris J.C., Heyman A., Mohs R.C. et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease // Neurology. – 1989. – Vol. 39, No. 9. – P. 1159–1165. – DOI: 10.1212/wnl.39.9.1159.

10. Armstrong T., Bull F. Development of the World Health Organization Global Physical Activity Questionnaire (GPAQ) // Journal of Public Health. – 2006. – Vol. 14, No. 2. – P. 66–70. – DOI: 10.1007/s10389-006-0024-x.

11. Bull F.C., Al-Ansari S.S., Biddle S. et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour // British Journal of Sports Medicine. – 2020. – Vol. 54, No. 24. – P. 1451–1462. – DOI: 10.1136/bjsports-2020-102955.

UDC 616.53-002:615.454.1

Aidarbek Zere Ramazanqyzy

12th grade student
NIS Almaty – Nauryzbay
(Almaty, Kazakhstan)

MECHANISMS OF ACTION AND BENEFITS OF SULFUR-CONTAINING PRODUCTS IN ACNE-MANAGEMENT

Annotation: Acne vulgaris appears to be one of the most prevalent inflammatory skin diseases among adolescents and adults around the world. Despite the huge amount of various dermatological and pharmaceutical formulations, many people prefer using older, simpler components such as sulfur. Sulfur's precise mechanisms of action and its role in maintaining skin's healthy condition have been discovered only in recent decades through research on its interaction with keratinocytes in epidermis and its conversion into bioactive hydrogen sulfide (H_2S). Nevertheless, this agent has been actively used in dermatology for many centuries. This article explores the mechanisms through which sulfur-containing formulations affect acne-prone skin (keratolytic, antimicrobial, anti-inflammatory, and antioxidant effects) and summarizes the available data about their benefits in using for acne management. Although empirical clinical evidence specific to acne vulgaris is still limited, existing knowledge suggests that sulfur can be considered as an applicable and generally well-tolerated agent for acne treatment or as an adjunct to other products.

Keywords: sulfur, acne vulgaris, hydrogen sulfide, keratolytic effect, acne treatment, sulfur-containing formulation

Introduction. Nowadays acne vulgaris affects a large proportion of young people, making them seek for effective dermatological care. Acne vulgaris is a multifactorial disease that involves excessive sebum production, abnormal keratinization of the follicular epithelium, proliferation of *Cutibacterium acnes,* and inflammation. Among current anti-acne formulations, retinoids, benzoyl peroxide, topical and systemic antibiotics, and hormonal therapies are the most common ones; however, these products are not always well tolerated by people with different skin types and might lead to irritation or other side effects.

Since antiquity sulfur, being reported by documented medicinal use, appears to be one of the oldest agents that are used in dermatology [1]. Thanks to its positive effects on skin's upper layers, sulfur has been used to mitigate a wide range of skin conditions, such as acne, rosacea, seborrheic dermatitis, and scabies, throughout many centuries [1]. However, despite the fact that sulfur has this long history of practical usage, there was not enough evidence about its mechanism of action, and the molecular basis of sulfur's effects on skin has been clarified relatively recently. This was achieved via research on the reaction that takes place when sulfur interacts with cysteine - a sulfur-containing amino acid that is highly abundant in skin proteins - in keratinocytes, leading to generation of cystine and hydrogen sulfide, which has a wide range of positive effects on epidermis [1].

This article focuses specifically on acne vulgaris and aims to explain the ways in which sulfur-containing formulations act on acne-prone skin and what benefits they have in acne-treatment therapies.

Sulfur's Mechanisms of Action on the Skin

Keratolytic effect. Topically applied sulfur undergoes chemical reaction with cysteine, leading to the formation of dimer called cystine and hydrogen sulfide (H_2S) as a byproduct (Figure 1). Produced cystine consists of two molecules of cysteine connected with a disulfide bond, which cross-links keratin proteins and increases their stability and strength. Applied in lower concentrations, sulfur tends to maintain normal keratinization, a process of keratinocytes' cytodifferentiation. On the other hand, high concentration of sulfur in dermatological formulations might lead to the accumulation of hydrogen sulfide, which is believed to cleavage bonds between keratin filaments in skin. As a result, under the effect of accumulated hydrogen sulfide, the outer skin layer can get softened and as dead skin cells are removed gradually shed [2]. This keratolytic effect appears to be clinically relevant and practically efficient for patients with acne-prone skin since it helps with removal of dead skin, which can get stuck in skin pores and contribute to the formation of comedones.

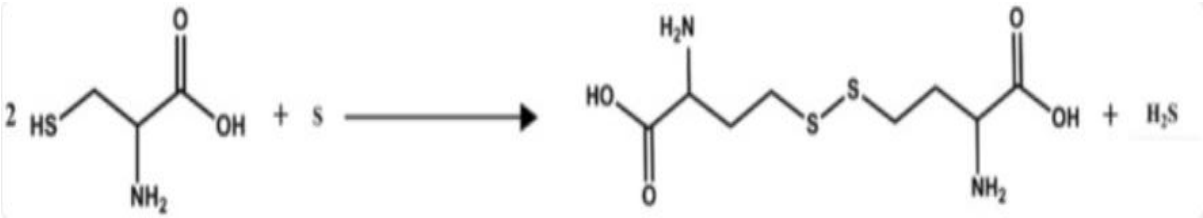


Figure 1. Chemical interaction between sulfur (S) and cysteine content of keratinocytes and formation of cystine and hydrogen sulfide as a byproduct.

Antimicrobial activity. Acne vulgaris is often caused by the proliferation of bacteria called *Cutibacterium acnes* in clogged follicles. Beyond an overall antibacterial effect that sulfur has on skin, this agent can be converted into compounds, for example pentathionic acid ($H_2O_6S_5$), which has been described as toxic to some certain microorganisms. In addition, sulfide salts tend to interact with microbial enzymes by causing protein precipitation and denaturation [1]. In a recent research study, it has been shown that formulations containing sulfur nanoparticles might have an antimicrobial activity against staphylococcal strains, which were collected from acne lesions and are more likely to carry resistance genes [3]. This finding suggests that new cosmetic products that contain precipitated sulfur can be produced in order to mitigate problems with antibiotic resistance.

Anti-inflammatory effects. Alongside with clogged pores, inflammation is considered as the main reason for all acne pustules and papules. Hydrogen sulfide, produced during the reaction between sulfur and cysteine as a byproduct, is a gaseous signaling molecule in the human body. It can help control or reduce inflammation by modulating several pathways that cause inflammatory processes to occur. Empirical evidence indicates that hydrogen sulfide, acting as a signaling molecule, tends to interact with key inflammatory pathways, including NF-κB and Nrf2. The mitigation of inflammatory immune response is achieved through inhibition of T-cell proliferation and production of cytokines such as IL-2, TNF- α, IL-6, and IL-8 [2]. Using keratinocyte cell models, scientists via TNF- α artificially increased inflammation and identified that hydrogen sulfide can reduce the amount of nitric oxide (NO), IL-6, and IL-8,

which are molecules taking part in skin inflammation processes [4]. These anti-inflammatory properties of hydrogen sulfide, a sulfur's derivative compound, are beneficial for acne management, especially to combat inflamed acne regions such as papules and pustules.

Antioxidant properties. Hydrogen sulfide derived from sulfur is more likely to exhibit antioxidant activity and, thereby, protect skin from getting damaged by harmful molecules. This antioxidant effect is reached through reduced glutathione, which acts as a protective shield of the skin and neutralizes oxidative substances, and direct scavenging of reactive oxygen species (ROS) such as superoxide anions (O_2^-) and hydrogen peroxide (H_2O_2) that can damage cells and lead to their oxidation [2]. Although this mechanism of sulfur derivative's action is often considered in cases of aging, UV damage, and oxidative stress experienced by skin, it appears to be helpful during the production of anti-acne formulations in order to constrain inflammatory cycle, often maintained by reactive oxygen species (ROS), in acne-prone skin.

The Use of Sulfur-Containing Cosmetic Formulations in Acne-Management. Sulfur is added to cosmetic and dermatological products mainly in two forms: precipitated and sublimed. In comparison to sublimed sulfur, precipitated one has a smaller particle size and greater surface area, which means that this form of sulfur tends to spread more easily in formulations and better interact with the skin surface due to its physical properties. Hence, precipitated sulfur is generally considered to be therapeutically superior and has better skincare and medical effects. Both forms - sublimed and precipitated - appear in sulfur soaps, face cleansers, lotions, gels, and face creams [1].

Examined specifically for acne treatment, sulfur concentrations in products that can be applied topically tend to range from low percentages (1-5%) up to around 10%. In order to improve the holistic effect, sulfur is often combined with other active ingredients. For example, the combination of sodium sulfacetamide (10%) with sulfur (5%) appears to be the one of the most studied combinations in acne-management [5]. While sodium sulfacetamide acts as antimicrobial agent helping with bacteria and inflammation, sulfur can be considered as an agent against oiliness, microorganisms, and dead skin cell buildup in pores. This combination is used in formulations such as creams, cleansers, or foams, and can be utilized to address not only acne vulgaris but also other skin conditions such as rosacea and seborrheic dermatitis [5]. Sulfur has also been found to be present in skin products combined with hydrocortisone, a mild corticosteroid, since this ingredient is likely to reduce inflammation and decrease redness and irritation [6]. Despite the limit of scientific evidence on efficacy, some stronger preparations containing sublimed sulfur (around 20%) are still in use for acne vulgaris [1].

Clinical Effectiveness of Sulfur in Acne Treatment. Although there are many newly formulated acne treatments in a modern clinical practice, sulfur continues to be actively used by dermatologists, especially for pustular acne [1]. This compound is considered as an effective agent among dermatologists due to its various effects: starting from keratolytic mechanism, which occurs through the damage of disulfide bonds between corneocytes under the hydrogen sulfide exposure, and ending with antimicrobial properties [2]. To evaluate the medical efficacy of sulfur, a Cochrane systematic review was performed by researches, which appears to be one of the most qualitative methods to identify the compound's role in clinical practice [7]. Examining several formulations for acne-prone skin, the findings indicate that sulfur was included in the review alongside with azelaic acid, salicylic acid, nicotinamide, zinc, and alpha-hydroxy acids that are considered as non-antibiotic topical treatments [7]. Based on these results, it can be concluded that although sulfur is much older than other current medications, it

continues to be clinically relevant for acne management. However, in comparison to benzoyl peroxide and retinoids, which are extensively studied through clinical investigations, sulfur remains relatively less examined, and there is limited amount of direct trial data, where its specific effects in acne treatment would be thoroughly explored.

Moreover, beyond traditional formulations – such as creams, soaps, or cleansers – sulfur nanoparticles, examined during laboratory studies, have demonstrated high antimicrobial activity: because of their extremely tiny size, these particles tend to interact with bacteria more efficiently and inhibit their proliferation [3]. For instance, it has been shown that sulfur nanoparticles might act against acne-related bacteria, including methicillin-resistant staphylococcal species (e.g. *Staphylococcus aureus*), which have developed resistance to antibiotics [3]. This antibacterial activity underlines sulfur-based therapies' role in acne-management and serves as a base for future practical implications, especially in addressing antibiotic-resistant, acne-related bacteria [3].

Safety and Tolerability. Generally, sulfur-based topical formulations are considered to be well-tolerated and not causing any severe adverse reactions; however, some mild side effects might still occur. An unpleasant odor and skin dryness (xerotic eczema) appear to be the most frequently reported issues associated with sulfur [8]. In fact, sulfur is likely to have a characteristic smell, sometimes described as a smell of rotten eggs since sulfur can produce hydrogen sulfide. Although this gas is not harmful to the organism, many people find it unpleasant. Dry skin can be stated as a side effect because of sulfur's keratolytic activity, which causes dead skin cells to shed, and its ability to remove an excessive amount of skin sebum. Other reported side effects include local irritation, which is usually temporal, and development of an allergic reaction to sulfur or another ingredient in the product applied [9]. Overall, topical formulations that contain sulfur at concentration of 6% or below are considered potentially safe for a wide range of people: children older than 2 months, pregnant women, and those who do breastfeeding [8]. However, because skin types and sensitivity vary across all individuals, it is recommended to do patch testing for people with highly-reactive skin before applying the product on a whole face.

Conclusion. Sulfur-containing cosmetic products act on acne-prone skin through a combination of keratolytic, antimicrobial, anti-inflammatory, and antioxidant mechanisms. These effects are mainly caused by sulfur's chemical conversion into hydrogen sulfide and other effective derivatives during the interaction with skin cells. While this compound is considered as well tolerated by majority of people and has a significant role in acne-management as a non-antibiotic topical agent, comprehensive clinical trials on its specific efficacy and mechanisms of action remain scarce. Future research studies might focus on sulfur alone without any additives in order to better define the optimal concentrations, formulation contents, and further sulfur-based therapies that would be beneficial for patients with acne-prone skin and acne vulgaris.

References:

1. Lin A. N. Sulfur Revisited / A. N. Lin, R. J. Reimer, D. M. Carter // Journal of
2. the American Academy of Dermatology. – 1988. – Vol. 18, No. 3. – P. 553–558.
3. Xiao Q. Hydrogen Sulfide in Skin Diseases: A Novel Mediator and Therapeutic Target / Q. Xiao, L. Xiong, J. Tang, L. Li, L. Li // Oxidative Medicine and Cellular Longevity. – 2021. – Vol. 2021. – e6652086.

4. Hashem N. M. Antimicrobial Activities Encountered by Sulfur Nanoparticles Combating Staphylococcal Species Harboring sccmecA Recovered From Acne Vulgaris / N. M. Hashem, A. Hosny, A. A. Abdelrahman, S. Zakeer // AIMS Microbiology. – 2021. – Vol. 7, No. 4. – P. 481–498.

5. Huang A. The Use of Balneotherapy in Dermatology / A. Huang, S. Seité, T. Adar // Clinics in Dermatology. – 2018. – Vol. 36, No. 3. – P. 363–368.

6. Draelos Z. D. The Multifunctionality of 10% Sodium Sulfacetamide, 5% Sulfur Emollient Foam in the Treatment of Inflammatory Facial Dermatoses / Z. D. Draelos // Journal of Drugs in Dermatology. – 2010. – Vol. 9, No. 3. – P. 234–236.

7. Harlan S. L. Steroid Acne and Rebound Phenomenon / S. L. Harlan // Journal of Drugs in Dermatology. – 2008. – Vol. 7, No. 6. – P. 547–550.

8. Liu H. Topical Azelaic Acid, Salicylic Acid, Nicotinamide, Sulphur, Zinc and Fruit Acid (Alpha-Hydroxy Acid) for Acne / H. Liu, H. Yu, J. Xia [et al.] // Cochrane Database of Systematic Reviews. – 2020. – Vol. 5, No. 5. – CD011368.

9. Uzun S. Clinical Practice Guidelines for the Diagnosis and Treatment of Scabies / S. Uzun, M. Durdu, A. Yürekli [et al.] // International Journal of Dermatology. – 2024. – Vol. 63, No. 12. – P. 1642–1656.

10. Lee C. C. Sulfur Spring Dermatitis / C. C. Lee, Y. H. Wu // Cutis. – 2014. – Vol. 94, No. 5. – P. 223–225.

11. Chen Y. J. Sulfur and Its Derivatives in Dermatology: Insights Into Therapeutic Applications—A Narrative Review / Y. J. Chen, J. Tang, L. Wang, W. Hua // Journal of Cosmetic Dermatology. – 2025. – Vol. 24, No. 8. – e70402.

UDC 616-006.6:004.85

Ibraimkhan Alikhan

Grade 10 student
Scientific supervisor: Bilimkhan Nazerke
Bilim Innovation Lyceum for Gifted Boys
(Karagandy, Kazakhstan)

A TWO-GENE CLASSIFIER SEPARATES BREAST TUMOR FROM NORMAL TISSUE ALMOST PERFECTLY WITHIN ONE COHORT, YET FAILS TO CARRY ITS DECISIONS TO NEW COHORTS: A CAUTIONARY STUDY OF MINIMAL GENE PANELS

Abstract: Distinguishing tumor from normal tissue using gene expression is one of the most common machine-learning exercises in cancer bioinformatics, and published reports routinely achieve near-perfect accuracy. We asked three linked questions in breast cancer (BRCA): how few genes are actually needed to separate tumor from normal tissue, whether a simple interpretable model can match a black-box model, and whether a panel learned in one cohort still works in independent cohorts. Using 1,211 samples from The Cancer Genome Atlas (TCGA), we found that discrimination saturates almost immediately: a single gene reaches a cross-validated area under the ROC curve (AUC) of 0.99, and a two-gene panel (SPRY2, FIGF) reaches 0.996. A logistic regression matched random forests and gradient boosting on held-out data (the logistic-versus-forest AUC differed by less than 0.001, 95% confidence interval -0.002 to 0.000), so model complexity added nothing. The important result is what happened next. When we applied the TCGA-trained panel to two independent cohorts measured on different microarray platforms and drawn from different populations (Irish and Malaysian), discrimination degraded but survived (AUC 0.97 and 0.82 for a stability-selected consensus panel), whereas the model's decision threshold did not transfer at all. On the most distant cohort, balanced accuracy at the inherited threshold fell to 0.50, no better than chance, and was recovered only when the threshold was re-tuned using that cohort's own labels. A larger, bootstrap-stable panel generalized better than the bare two-gene panel but did not repair calibration. We conclude that the near-perfect numbers reported for tumor-versus-normal classification reflect how separable the problem is within a single dataset, not how well a minimal panel would work in practice, and that honest evaluation of such panels must test discrimination and calibration separately across more than one cohort.

Keywords: breast cancer, gene expression, machine learning, minimal gene panel, external validation, calibration, bioinformatics

Introduction. Breast cancer is the most commonly diagnosed cancer in women worldwide [1], and distinguishing malignant from healthy tissue is central to both its diagnosis and its study. One way researchers describe a tissue sample is by its gene expression: how strongly each of roughly twenty thousand genes is switched on, measured across the whole sample at once. Because tumor cells behave very differently from healthy cells, their gene expression differs too, and a natural question is whether a computer can learn to tell the two apart from these measurements alone.

For breast cancer, the answer has been "yes" for more than a decade, and classifiers that separate tumor from normal breast tissue commonly report accuracies above 95% [2]. Throughout this paper we measure that ability with the area under the ROC curve (AUC), a standard score between 0.5 (no better than a coin flip) and 1.0 (a model that always ranks a tumor above a normal sample). Because the result is so reliably strong, the more interesting scientific questions are no longer whether the classification can be done, but how economically it can be done, how simply, and how durably: with how few genes, with how simple a model, and with how much confidence that the answer will hold on patients the model has never seen.

These three questions matter for different reasons. A small gene panel is cheaper to measure and easier to interpret biologically than a whole-transcriptome model, which is why much of the clinical translation of expression research has been the search for compact signatures [3]. A simple model such as a logistic regression or a shallow decision tree can be read and checked by a human, whereas a random forest or gradient-boosted ensemble cannot; if the simple model performs as well, the added opacity of the complex one is not worth paying for. Generalization across cohorts is the property that separates a real biomarker from an artifact of one dataset. The history of expression-based signatures includes many that looked excellent in the cohort that produced them and then failed to reproduce, often because of batch effects and population differences between datasets [4, 5, 6].

Two features of the standard tumor-versus-normal task make it an unusually clear setting in which to study these questions. First, the signal is large: tumor and normal tissue differ in the expression of thousands of genes, so the classification is easy and near-perfect scores are expected. Second, that very easiness is a trap. When every reasonable model scores close to the ceiling within one dataset, internal cross-validation can no longer distinguish a genuinely robust panel from a fragile one, because they all look identical. The only way to tell them apart is to leave the dataset.

We therefore designed this study not to build the most accurate possible classifier, a target that is already saturated and uninteresting on its own, but to ask what internal performance does and does not tell us. Concretely: (i) we measured how AUC rises with panel size in TCGA breast cancer to locate the smallest panel that captures the available signal; (ii) we compared interpretable models against black-box models on that panel; and (iii) we tested whether the panel and the trained model transfer to two independent cohorts on different platforms and from different populations, separating the question of whether the model can still rank samples (discrimination) from whether its actual decisions remain correct (calibration). Our hypothesis was that internal near-perfection would overstate external performance, and specifically that minimal panels chosen on internal data would be optimistic. The data supported a sharper version of this hypothesis than we had expected.

Materials and Methods. Data. Discovery data were the TCGA-BRCA "HiSeqV2" gene-level expression matrix obtained from the UCSC Xena platform [9], comprising log2-transformed, normalized RNA-seq values for 20,531 genes. Sample type was read from the TCGA barcode: codes ending in -01 were labeled primary tumor (n = 1,097) and -11 solid-tissue normal (n = 114); the seven metastatic (-06) samples were excluded to keep the contrast clean. Genes with missing values or near-zero variance were removed, leaving 20,250 genes. External cohorts were GSE42568 [10] (GPL570; 104 tumors, 17 normals) and GSE15852 [11] (GPL96; 43 tumors, 43 normals), both downloaded as GEO series matrices; probe identifiers were mapped to gene symbols using the corresponding GEO platform annotation and collapsed to

gene level by averaging probes. Because GEO series matrices are inconsistently scaled, any cohort whose 99th-percentile intensity exceeded 50 was log2-transformed so that all cohorts shared a log scale.

Feature selection and panel definition. All gene selection used only training data. Within the TCGA training set (70%, stratified), genes were ranked by the absolute value of a Welch two-sample t-statistic between tumor and normal. The panel-size curve (Figure 1) was computed by nested 5-fold cross-validation: in each fold the top k genes were reselected on that fold's training portion and a logistic regression was scored on the held-out portion, so the curve is not contaminated by the validation fold. The minimal panel was the smallest k whose mean cross-validated AUC was within 0.005 of the maximum. Panel stability was assessed two ways: the mean pairwise Jaccard overlap of the size-k panels selected across the five folds, and the selection frequency of each gene among the top 20 across 500 bootstrap resamples of the training set, which defined the bootstrap-stable consensus ordering.

Models and evaluation. On the locked panel we trained logistic regression (L2-regularized), a depth-3 decision tree, a random forest (400 trees), and gradient boosting, all with balanced class weights to address the approximately 10:1 class imbalance. Features were standardized using statistics estimated on the training set only. Each model was scored once on the untouched 30% TCGA test set; we report AUC, average precision (PR-AUC), accuracy, balanced accuracy, F1, sensitivity, and specificity. The logistic-versus-random-forest AUC difference and its 95% confidence interval were estimated by 2,000 bootstrap resamples of the test set. For external validation, models were trained on the full TCGA cohort (panel genes, z-scored within TCGA) and applied to each external cohort (same genes, z-scored within that cohort); external labels were never used for fitting. We report external AUC with a 2,000-resample bootstrap confidence interval, and we report balanced accuracy both at the transferred 0.5 decision threshold and at the cohort-optimal threshold (the point maximizing Youden's J on that cohort) to separate discrimination from calibration.

Reproducibility. Analyses used Python 3.14 with NumPy, pandas, SciPy, and scikit-learn [12]. A fixed random seed (20260613) was used throughout. All data are public; the analysis scripts, the exact dataset identifiers, and the result and figure files needed to reproduce every number in this paper are provided in the replication package.

Ethics. All data are de-identified, publicly released human genomic datasets used under their public data-use terms; no new human-subjects data were collected, and no attempt was made to re-identify any individual.

Results

Discrimination saturates with one to two genes

We split the 1,211 TCGA-BRCA samples (1,097 primary tumors, 114 solid-tissue normals) into a 70% training set and a 30% held-out test set, ranked genes by their tumor-versus-normal t-statistic within the training folds only, and measured how cross-validated AUC changed as we added genes (Figure 1). A single top-ranked gene already achieved a mean cross-validated AUC of 0.990; two genes reached 0.996; and performance plateaued near 0.999 by ten genes, with no meaningful improvement from adding more. By our pre-specified rule (the smallest panel within 0.005 AUC of the best), the minimal panel contained just two genes, SPRY2 and FIGF. On the untouched test set this two-gene logistic-regression panel scored an AUC of 0.9994.

This confirms, with a clear quantitative shape, the expectation that tumor-versus-normal discrimination in bulk breast tissue is an easy problem: essentially all of the separable signal is captured by the first one or two genes.

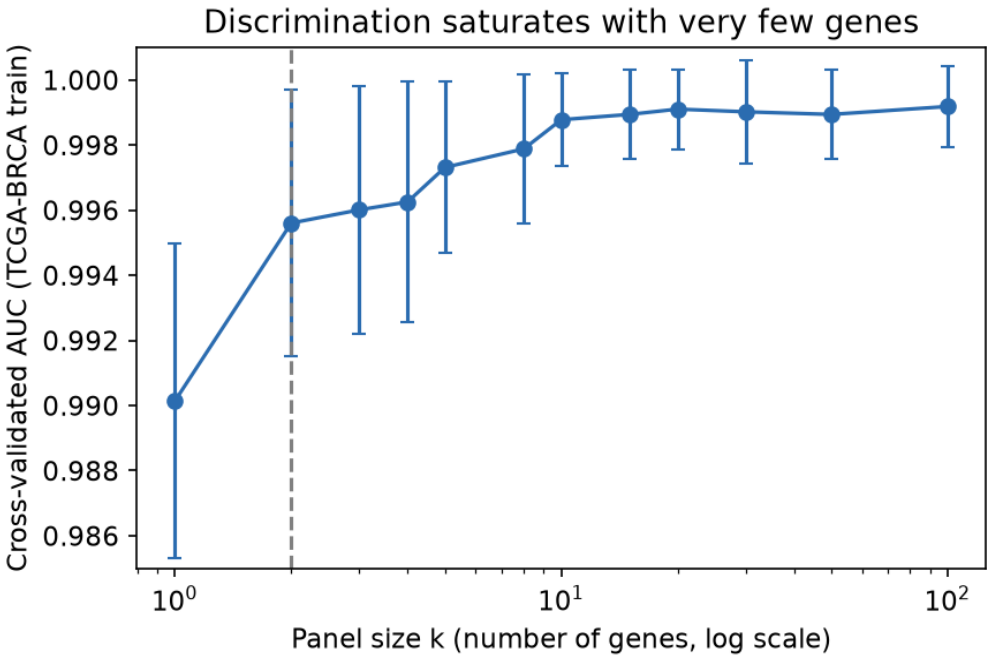


Figure 1. Discrimination saturates with very few genes. Mean cross-validated AUC (±SD across 5 folds) on the TCGA-BRCA training set as a function of panel size k (log scale). The dashed line marks the selected minimal panel (k = 2).

Simple models match black-box models

On the locked two-gene panel we trained four classifiers and scored each once on the held-out test set (Figure 2). A logistic regression (AUC 0.9994) and a depth-3 decision tree (AUC 0.9984), both fully interpretable, were statistically indistinguishable from a random forest (AUC 1.000) and gradient boosting (AUC 1.000). A bootstrap estimate of the AUC difference between logistic regression and random forest was -0.0006, with a 95% confidence interval of -0.0021 to 0.000. A random forest trained on all 20,250 genes did no better (AUC 0.9999) than the two-gene logistic model. These near-perfect scores reflect how separable the task is rather than information leakage: gene ranking and model fitting used the training data only, and the test set was scored exactly once. For this task, then, neither additional genes nor additional model complexity bought any measurable accuracy, and the interpretable model is preferable on every other ground.

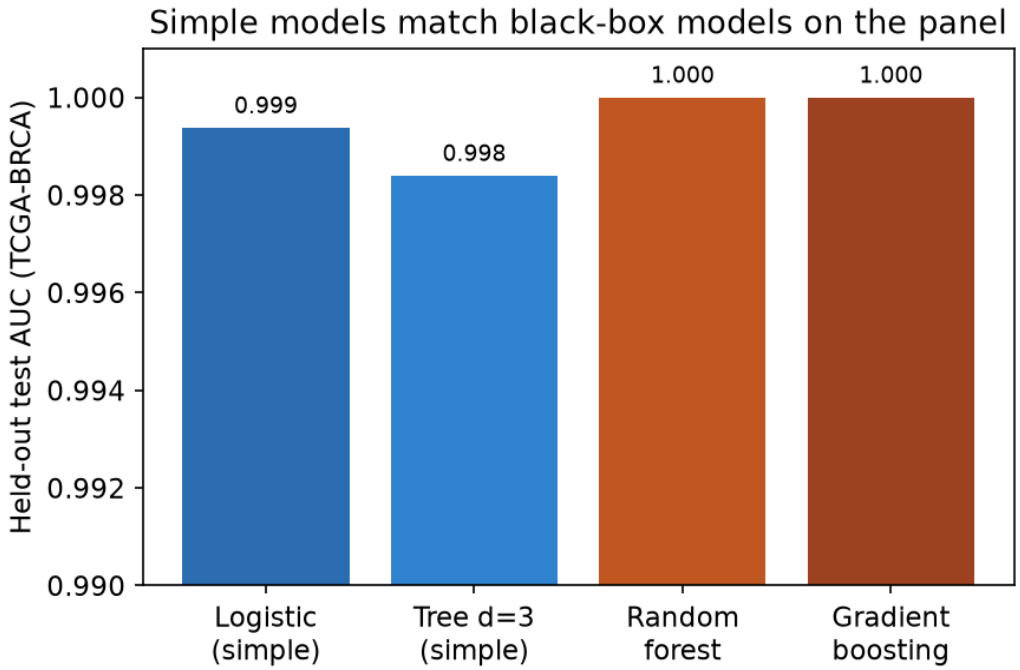


Figure 2. Simple models match black-box models on the panel. Held-out TCGA test AUC for the two-gene panel under a logistic regression, a depth-3 tree, a random forest, and gradient boosting.

The minimal panel is only moderately unique, but a stable consensus exists

Because so many genes are strongly differentially expressed, the particular two genes selected depend somewhat on the sample. Across five cross-validation folds the exact two-gene panel had a mean pairwise Jaccard overlap of 0.73, and in a 500-iteration bootstrap the second panel member was replaced in a substantial minority of resamples. Rather than treat one fragile pair as "the" panel, we ranked genes by how often they appeared among the top genes across 500 bootstraps (Figure 3). FIGF and SPRY2 were selected in 100% of bootstraps, followed by TSLP (99.8%), SCN4B (99.6%), TMEM220 (99.6%), MATN2 (99.6%) and MMP11 (98.0%). We carried forward this bootstrap-stable ordering as a "consensus panel" so that the genes we tested externally were not an artifact of a single split (Table 1). In the locked two-gene logistic model both genes were lower in tumor than in normal tissue (SPRY2 log2 fold-change -2.9; FIGF -6.1), consistent with the reported loss of the SPRY2 tumor-suppressor in breast cancer [7].

Table 1 – The bootstrap-stable consensus panel. Selection frequency is the fraction of 500 bootstrap resamples of the TCGA training set in which the gene appeared among the top 20 by t-statistic. Direction is the change in tumor relative to normal tissue.

Gene	Selection frequency	Direction in tumor
FIGF (VEGFD)	1.00	down
SPRY2	1.00	down
TSLP	0.998	down
SCN4B	0.996	down
TMEM220	0.996	down
MATN2	0.996	down
MMP11	0.980	up
MAMDC2	0.960	down

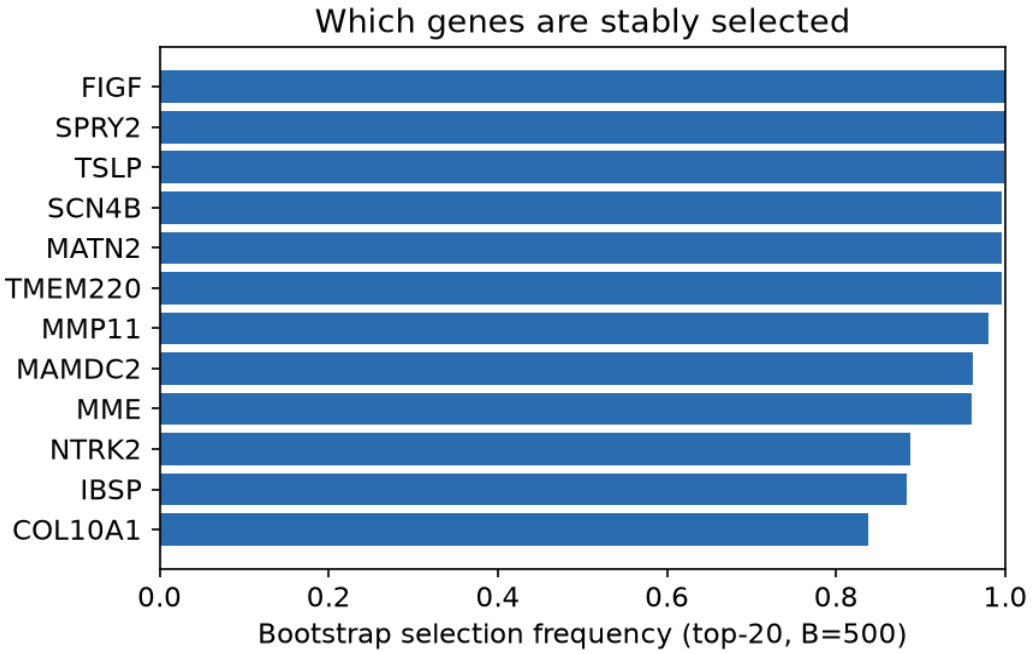


Figure 3. Which genes are stably selected. Bootstrap selection frequency (top-20 membership across 500 resamples of the TCGA training set) for the most frequently chosen genes.

Discrimination degrades gracefully across cohorts, but calibration does not transfer

We then applied the TCGA-trained panels, unchanged, to two independent cohorts (Table 2): GSE42568 (104 tumors, 17 normals; Irish; Affymetrix GPL570) and GSE15852 (43 tumors, 43 normals, matched; Malaysian; Affymetrix GPL96). Each cohort differs from TCGA in both measurement platform and population, and neither cohort's labels were ever used to fit or tune the model; each gene was standardized within each cohort to put the RNA-seq and microarray scales on comparable footing (Figures 4–6).

Table 2 – The three cohorts. The TCGA discovery cohort and the two independent validation cohorts differ in measurement platform and population, which is what makes them a genuine external test.

Cohort	Role	Platform	Population	Tumor	Normal
TCGA-BRCA	discovery	RNA-seq (Illumina HiSeq)	mixed (USA)	1,097	114
GSE42568	external	microarray (Affymetrix GPL570)	Irish	104	17
GSE15852	external	microarray (Affymetrix GPL96)	Malaysian	43	43

Discrimination held up reasonably well. The consensus panel reached an external AUC of 0.967 (95% CI 0.91–1.00) on the Irish cohort and 0.817 (95% CI 0.73–0.90) on the more distant Malaysian cohort; the bare two-gene panel was weaker on both (0.893 and 0.706). In every cohort the interpretable logistic model again matched or beat the random forest, so the simple-equals-complex finding was not an artifact of the discovery cohort.

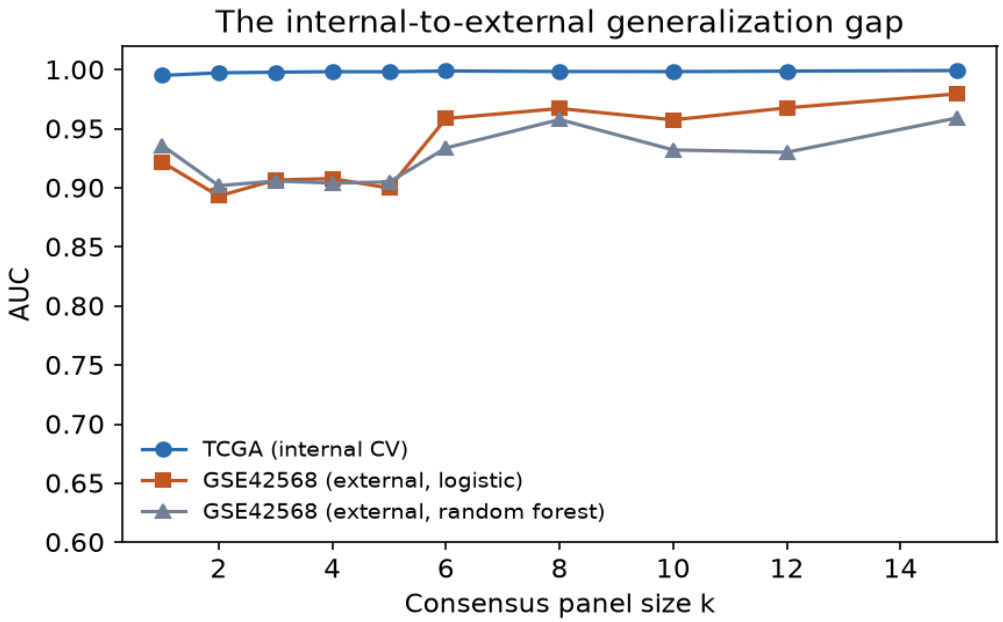


Figure 4. The internal-to-external generalization gap. AUC of the bootstrap-consensus panel as a function of size, evaluated internally (TCGA cross-validation) and externally on GSE42568 with logistic regression and random forest. Internal AUC is flat near the ceiling while external AUC is lower and size-dependent.

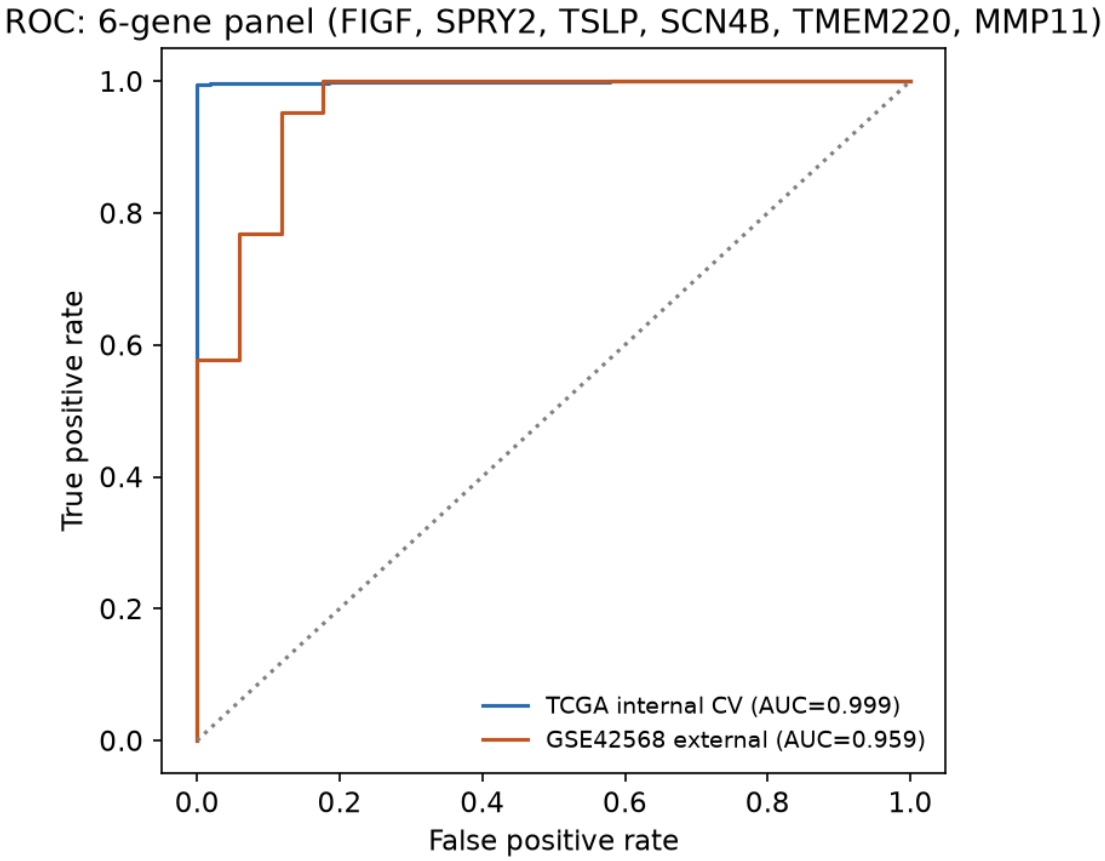


Figure 5. Internal versus external ROC. ROC curves for the consensus panel: TCGA internal cross-validation versus the GSE42568 external cohort.

The decisions, however, did not transfer. When we kept the threshold the model had learned on TCGA, balanced accuracy on the Irish cohort stayed high (0.91), but on the

Malaysian cohort it collapsed to 0.50 for the consensus panel, exactly chance, because the model labeled almost every sample as tumor (specificity 0.00–0.16). This was not a failure of the underlying signal: when we instead chose the threshold that was optimal on that cohort, balanced accuracy recovered to 0.76 (consensus panel) and 0.66 (two-gene panel). This recalibrated figure uses the external cohort's own labels to pick the threshold, so it is an optimistic best case rather than a deployable accuracy; we report it only to show that the ranking still carries real signal that the inherited threshold fails to use. The gap between the transferred-threshold and recalibrated-threshold balanced accuracy isolates the problem precisely. The model can still rank Malaysian samples well above chance, but the operating point it learned from TCGA is wrong for that cohort. On the Irish cohort, which is heavily tumor-weighted like TCGA, the transferred and recalibrated thresholds nearly coincided (0.91 vs 0.91), which is why the calibration failure is invisible there and only appears on the balanced, more distant cohort.

The larger consensus panel generalized better than the two-gene panel on discrimination in every external comparison, but it did not repair calibration; its transferred-threshold balanced accuracy was the one that fell furthest. More genes bought a more reliable ranking, not a more transferable decision rule.

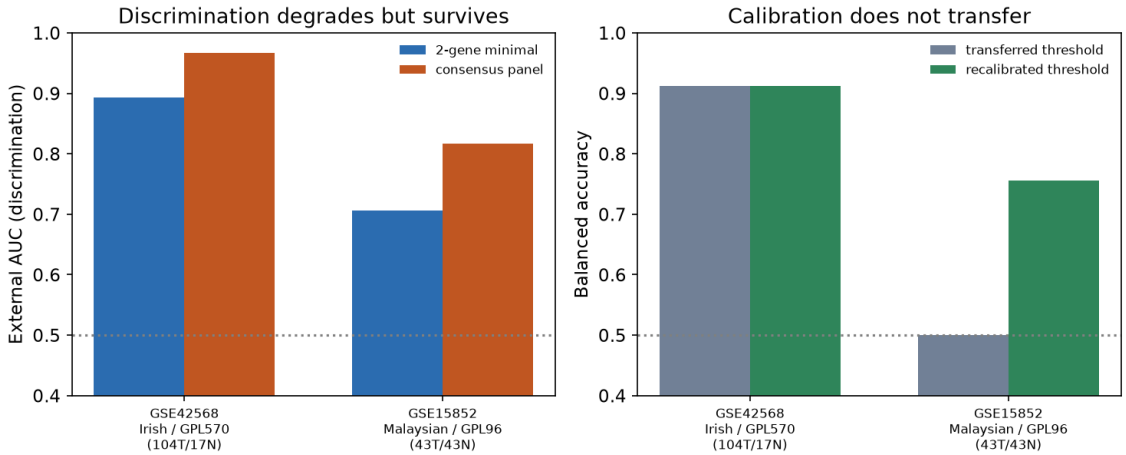


Figure 6. Discrimination degrades but calibration does not transfer. Left: external AUC of the two-gene and consensus panels on the Irish (GSE42568) and Malaysian (GSE15852) cohorts. Right: balanced accuracy of the consensus panel at the transferred TCGA threshold versus a cohort-recalibrated threshold; the transferred threshold collapses to chance on the most distant cohort.

Discussion. The headline number for tumor-versus-normal classification in breast cancer, near-perfect internal accuracy, is real and reproducible, and on its own almost meaningless as evidence that a minimal panel would be useful. Our internal results reproduce the saturation that the literature would predict: one gene is nearly enough, two genes essentially exhaust the separable signal, and a transparent logistic regression matches the best black-box model. If the study had stopped at the edge of the discovery cohort, as many do, it would have reported a tidy two-gene, fully interpretable, "100% AUC" classifier.

The external analysis shows why that report would have been misleading in a specific and measurable way. We found it useful to separate two things that internal cross-validation silently merges. Discrimination, the ability to rank tumors above normals, degraded gracefully across platform and population (AUC from approximately 0.99 internally to 0.82 on the most distant cohort) and remained clearly above chance; this is the part of performance that internal

optimism overstates only mildly. Calibration, whether the model's actual yes/no decisions remain correct at a fixed threshold, did not transfer at all to the most distant cohort, where the inherited threshold produced chance-level balanced accuracy despite a still-informative ranking. The distinction matters because a classifier is deployed at a threshold, not as a ranking, and a model that "still has an AUC of 0.82" can nonetheless make almost entirely wrong calls on a new population; calibration is increasingly recognized as the property of a clinical prediction model that most often fails on transfer, even when discrimination is preserved [8]. The widening gap between transferred and recalibrated balanced accuracy, from negligible on the TCGA-like Irish cohort to large on the Malaysian cohort, is a compact way to see external optimism that a single AUC hides.

Two design choices were necessary to see this. First, using two external cohorts rather than one separated a platform effect from a population effect well enough to show that the calibration problem grows with distance from the training cohort, not with any single dataset's quirk. Second, reporting balanced accuracy at both the transferred and the cohort-optimal threshold, rather than AUC alone, was what exposed the calibration failure; AUC by itself would have told a falsely reassuring story.

For a student or researcher building a minimal expression panel, the practical lessons are concrete. The size of a panel chosen by internal performance is not trustworthy on its own, because internal cross-validation cannot distinguish panels that an external cohort separates clearly; a bootstrap-stable consensus panel is a safer object to carry forward than the single best split's selection. Reported accuracy should always be paired with at least one external cohort and with a calibration check, not only a discrimination metric. And the common practice of stopping at internal validation will tend to publish panels that look deployable and are not.

This study has clear limitations. It examines a single, deliberately easy task (bulk tumor versus normal tissue), and the conclusions about calibration transfer may differ for harder problems such as subtype classification or outcome prediction, where the internal ceiling is lower and the questions of panel size and model complexity are live rather than saturated. The normal-tissue counts in the external cohorts are small (17 in one cohort), which widens the confidence intervals and is why we report them. The platform difference (RNA-seq versus two microarray generations) is entangled with the population difference, so we cannot fully separate the two sources of the calibration shift, only show that it grows with combined distance. Finally, the biological interpretation of the panel genes is secondary to the methodological point and is offered only as a consistency check, not as a claim of novel biology. Future work that we think would be valuable includes repeating the discrimination-versus-calibration analysis on a non-saturated task, and testing whether standard cross-dataset normalization methods restore calibration as well as they preserve discrimination.

Conclusion. In breast cancer, tumor-versus-normal discrimination from gene expression is an easy problem: one or two genes (SPRY2, FIGF) reach near-perfect internal AUC, and a transparent logistic regression matches black-box models. A panel selected on internal data, however, overstates its real-world usefulness. Across two independent cohorts on different platforms and populations, discrimination degraded gracefully (AUC down to 0.82) while the inherited decision threshold did not transfer at all, collapsing to chance-level balanced accuracy on the most distant cohort; a bootstrap-stable consensus panel improved discrimination but not calibration. Honest evaluation of minimal gene panels must therefore test discrimination and

calibration separately on more than one cohort, rather than relying on near-perfect single-dataset accuracy.

Acknowledgments. I thank Nazerke Bilimkhan for supervising this project. All datasets used are publicly available from The Cancer Genome Atlas (via UCSC Xena) and the NCBI Gene Expression Omnibus.

References:

1. Sung H., Ferlay J., Siegel R.L. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries // CA: A Cancer Journal for Clinicians. – 2021. – Vol. 71, No. 3. – P. 209–249. – DOI: 10.3322/caac.21660.

2. Sotiriou C., Pusztai L. Gene-expression signatures in breast cancer // New England Journal of Medicine. – 2009. – Vol. 360, No. 8. – P. 790–800. – DOI: 10.1056/NEJMra0801289.

3. Sørlie T., Perou C.M., Tibshirani R. et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications // Proceedings of the National Academy of Sciences. – 2001. – Vol. 98, No. 19. – P. 10869–10874. – DOI: 10.1073/pnas.191367098.

4. Leek J.T., Scharpf R.B., Bravo H.C. et al. Tackling the widespread and critical impact of batch effects in high-throughput data // Nature Reviews Genetics. – 2010. – Vol. 11, No. 10. – P. 733–739. – DOI: 10.1038/nrg2825.

5. Subramanian J., Simon R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? // Journal of the National Cancer Institute. – 2010. – Vol. 102, No. 7. – P. 464–474. – DOI: 10.1093/jnci/djq025.

6. Saeys Y., Inza I., Larrañaga P. A review of feature selection techniques in bioinformatics // Bioinformatics. – 2007. – Vol. 23, No. 19. – P. 2507–2517. – DOI: 10.1093/bioinformatics/btm344.

7. Lo T.L., Yusoff P., Fong C.W. et al. The Ras/mitogen-activated protein kinase pathway inhibitor and likely tumor suppressor proteins, sprouty 1 and sprouty 2 are deregulated in breast cancer // Cancer Research. – 2004. – Vol. 64, No. 17. – P. 6127–6136. – DOI: 10.1158/0008-5472.CAN-04-1207.

8. Van Calster B., McLernon D.J., van Smeden M. et al. Calibration: the Achilles heel of predictive analytics // BMC Medicine. – 2019. – Vol. 17. – Art. 230. – DOI: 10.1186/s12916-019-1466-7.

9. Goldman M.J., Craft B., Hastie M. et al. Visualizing and interpreting cancer genomics data via the Xena platform // Nature Biotechnology. – 2020. – Vol. 38, No. 6. – P. 675–678. – DOI: 10.1038/s41587-020-0546-8.

10. Clarke C., Madden S.F., Doolan P. et al. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis // Carcinogenesis. – 2013. – Vol. 34, No. 10. – P. 2300–2308. – DOI: 10.1093/carcin/bgt208.

11. Pau Ni I.B., Zakaria Z., Muhammad R. et al. Gene expression patterns distinguish breast carcinomas from normal breast tissues: the Malaysian context // Pathology – Research and Practice. – 2010. – Vol. 206, No. 4. – P. 223–228. – DOI: 10.1016/j.prp.2009.11.006.

12. Pedregosa F., Varoquaux G., Gramfort A. et al. Scikit-learn: machine learning in Python // Journal of Machine Learning Research. – 2011. – Vol. 12. – P. 2825–2830.

UDC 81'27

Komekova Saida

Grade 10 student
Scientific supervisor: Zhunusova Lida Seilbekqyzy
Binom School named after Qanysh Satbayev
(Astana, Kazakhstan)

TESTING CLASSIC CODE-SWITCHING CONSTRAINTS ON KAZAKH–RUSSIAN SOCIAL-MEDIA TEXT: AN INSERTIONAL CORPUS STUDY

Abstract: Classic structural accounts of intra-sentential code-switching, Poplack's Free-Morpheme Constraint (FMC) and Equivalence Constraint (EC), and Myers-Scotton's Matrix Language Frame (MLF) model, were developed largely on Spanish–English and other Indo-European pairs. How they behave on a typologically mismatched agglutinative–fusional pair such as Kazakh–Russian is under-described, especially in naturalistic written social-media data. This study assembles a corpus of 2,447 Kazakh-matrix sentences containing Russian-script material, drawn from a public, openly-licensed set of 172,015 Kazakh online reviews (KazSAnDRA, CC BY 4.0), and codes the structural type of each switch and, for single-item switches, whether the Free-Morpheme Constraint is respected. Kazakh–Russian online code-switching is overwhelmingly insertional: about 98% of switch spans are single words, and Kazakh is the matrix language in about 98% of sentences. The typological prediction that the Free-Morpheme Constraint would be the first to fail for this pair is not supported. Russian elements are mostly inserted as bare free morphemes, and on a cleaned set of 1,618 single-item lexical switches the FMC is violated in only 4.9% of cases (95% CI 3.9–6.0) and respected in 95.1%. The rate is stable across borrowing thresholds (4.7–5.1%) and barely moves between the raw and cleaned candidate sets. This contrasts with spoken findings on comparable pairs: in Turkish–English [1], Arabic–English [2], and, most tellingly, the closest sister pair Kyrgyz–Russian [3], morphological integration of inserted material is common, and in spoken Kyrgyz–Russian the single most frequent insertion pattern is a Russian noun carrying a native matrix-language suffix. Written Kazakh–Russian switching is just as insertional and just as firmly Kazakh-matrix as those pairs, but keeps the inserted Russian element morphologically bare, so the Free-Morpheme Constraint is respected far more often. The violations that do occur cluster in slang and recent loans that take Kazakh case or number morphology (for example мэмбер-лер-ге 'to the members', and a vulgar slang neologism carrying the Kazakh plural suffix -тар). In this written register, Kazakh–Russian contact is dominated by lexical insertion that mostly conforms to the classic constraints, with morphological integration a minority pattern. The most economical reading of the contrast with the spoken neighbours is a register-and-modality effect rather than a typological one. All data, code, and coding decisions are released for reproduction.

Keywords: code-switching; Kazakh–Russian; Free-Morpheme Constraint; Matrix Language Frame model; corpus linguistics; social-media language; bilingualism.

1. Introduction

Kazakhstan is a deeply bilingual country. Kazakh is the state language and Russian is the language of interethnic communication, and in cities most ethnic Kazakhs use both, often

moving between them inside a single sentence. This kind of switching, where a person writing or talking mainly in Kazakh drops in a Russian word or phrase, is an ordinary feature of everyday language in Almaty, Astana, and online. It appears in conversation, in messaging, and in the short evaluative writing people produce when they review a product, a film, or a service. For most urban Kazakh speakers it is not a marked or unusual act; it is simply how they communicate.

The pairing is interesting for linguistics because the two languages are built very differently. Kazakh is agglutinative: it adds strings of suffixes to a stem to mark case, number, and possession, so that grammatical information is spread across a sequence of clearly separable endings (for example *үй* 'house', *үй-ге* 'to the house', *үй-лер-ге* 'to the houses'). Russian is fusional: it packs several grammatical meanings into a single ending, so that one suffix can simultaneously carry case, number, and gender. When two languages this structurally distinct are mixed, the points where they join become a natural test of the rules that are supposed to govern code-switching. If a Russian noun is dropped into a Kazakh sentence, does the speaker leave it bare, or do they attach Kazakh suffixes to it as if it were a Kazakh word? The answer is not obvious in advance, and different theoretical traditions predict different things.

Three proposals have shaped how linguists think about the structure of switching. The Free-Morpheme Constraint and the Equivalence Constraint [4, 5] and the Matrix Language Frame model [6] were built mainly on Spanish–English and other European pairs. How they behave on a typologically mismatched, agglutinative–fusional pair like Kazakh–Russian is much less clear, and almost no work tests them on naturally written social-media text, where switching happens constantly and is easy to collect at scale. Most existing descriptions of Kazakh–Russian contact are functional or sociolinguistic: they describe who switches, in what settings, and why, rather than offering structural tests of whether the classic constraints hold.

This study asks a deliberately narrow question: in written Kazakh–Russian code-switching, do these constraints hold, and which one is broken most often? To answer it, the study assembles a corpus of 2,447 Kazakh-matrix sentences containing Russian material from a public dataset of Kazakh online reviews, codes the structure of each switch, and measures how often the Free-Morpheme Constraint is respected. The contribution is a first quantitative, reproducible, openly-licensed test of these classic constraints on this under-studied pair, in a written register that previous structural work on the closely related pairs has largely not examined. The result is reported with explicit attention to the borrowing problem, the genuine difficulty of telling an inserted code-switch apart from an established loanword, which is the single most important methodological hazard in this kind of count and which much earlier work leaves implicit.

2. Background and Literature

2.1 Code-switching, borrowing, and the unit of analysis

Code-switching is the use of two languages within a single conversation or sentence. *Inter-sentential* switching moves between languages at sentence boundaries; *intra-sentential* switching, the focus of this study, happens inside a single sentence, where one language supplies most of the clause and the other contributes one or more elements. Researchers have long asked whether intra-sentential switching is random or rule-governed, and the dominant answer has been that it is rule-governed: switches are not scattered arbitrarily but cluster at particular structural sites and avoid others.

A persistent complication runs underneath every structural count of switching: the difference between a code-switch and a borrowing. A code-switch is a momentary use of the other language; a borrowing is a foreign-origin word that has become an established part of the recipient language and is no longer felt as foreign. The distinction matters because an established borrowing that happens to carry native morphology is not evidence that a speaker violated a switching constraint; the word is simply a recipient-language word by now. Muysken [7] treats the boundary as a continuum rather than a clean line, organised by degree of bilingual use, regularity, and structural integration, and warns that any quantitative claim about switching depends heavily on where the analyst draws it. This study takes that warning seriously and builds an explicit, reproducible borrowing filter rather than leaving the decision to intuition (Section 4).

2.2 The three classic constraints

The Free-Morpheme Constraint (FMC) [4, 5] states that a switch cannot occur between a bound morpheme and a lexical stem unless the stem is phonologically integrated into the language of the morpheme. In plain terms, a speaker should not attach a grammatical suffix from one language directly onto a root from the other in the middle of a switch. For the present pair, the configuration the FMC forbids is a Russian root that carries a Kazakh case, number, or possessive suffix without having been adopted as a Kazakh word, for instance a Russian noun stem with a Kazakh dative or plural ending attached.

The Equivalence Constraint (EC) [4] states that switches tend to occur at points where the two languages share the same surface word order, and are avoided where their orders clash. The constraint is most naturally tested on multi-word switches and on the boundaries between switched and non-switched stretches, because it is fundamentally a claim about word order at the switch site. For a single inserted content word it is often not testable, since one word does not by itself create an order conflict.

The Matrix Language Frame model (MLF) [6] takes a different angle. In any mixed clause, one language, the *matrix*, supplies the grammatical frame, including the inflectional morphology and the order of major constituents, while the other, the *embedded* language, contributes content words inserted into that frame. The MLF predicts an asymmetry: system morphemes (the grammatical glue) come from the matrix, while embedded-language material is mostly open-class content. Identifying the matrix language is therefore the first analytic step, and for Kazakh–Russian the relevant question is whether Kazakh reliably supplies the inflectional frame.

2.3 Evidence from typologically distinct pairs

These constraints have not held up uniformly, and the clearest counter-evidence comes from typologically distinct pairs in which one language is agglutinating. For Arabic–English, Bahri [2] analysed 886 intra-sentential code-switching utterances from naturalistic podcast speech among Saudi women and found substantial violations of both constraints: the FMC was violated in 34.5% of utterances and the EC in 67.8%. The FMC violations were unidirectional (Arabic bound morphemes attaching to English free morphemes), and the same speakers alternated between constraint-abiding and constraint-violating utterances, which rules out limited competence as the explanation. For Turkish–English, an agglutinative–analytic pair, Kemaloğlu-Er [1] reported that the most salient intra-sentential pattern was the affixation of Turkish bound morphemes (case, plural, and possessive suffixes) to English noun stems with no phonological adaptation, a direct and frequent violation of the FMC; the data were only

partially consistent with the EC. Because Kazakh is also agglutinative, the natural prediction by analogy with Turkish–English and Arabic–English is that Kazakh–Russian should break the FMC readily, with Kazakh case and number suffixes attaching to inserted Russian roots.

The closer neighbours, Turkic- and Caucasus-and-Russian pairs, agree on one thing and disagree on another, and the disagreement is exactly what this study measures. They agree that switching is insertional and that the agglutinative language is the stable matrix. İnan [8] analysed Turkish–Russian code-switching among bilingual Ahıska Turks and found the data well-described by the MLF model across parts of speech, with Russian adpositions surfacing as embedded-language islands. Forker [9], working on Sanzhi (a Nakh-Daghestanian language) in contact with Russian, analysed roughly 6,000 tokens of natural speech from six speakers, found the data largely handled within the MLF model, confirming Myers-Scotton's Uniform Structure Principle, and reported a strong preference for inserting open-class items, especially nouns, over closed-class items. Where these studies bear directly on the Free-Morpheme Constraint, though, the spoken data show heavy morphological integration of the inserted item.

The sharpest case is the closest relative. Odagiri [3] analysed Kyrgyz–Russian switching in Bishkek, Kyrgyz being a Turkic sister of Kazakh, and found that the single most frequent insertion pattern of all was a Russian noun carrying a Kyrgyz (matrix-language) suffix, with the same morphological adaptation observed for inserted adjectives and verbs. That is, the configuration the FMC forbids is not marginal in spoken Kyrgyz–Russian; it is the dominant one. Odagiri also foregrounds the methodological problem this study runs into directly: the boundary between an inserted code-switch and an established borrowing is genuinely hard to draw, and is best treated as a continuum of bilingual use, regularity, and structural integration [7].

2.4 The gap and the competing expectations

The literature therefore offers two competing expectations for Kazakh–Russian. The spoken agglutinative-pair studies (Turkish–English, Arabic–English, and even the sister pair Kyrgyz–Russian) predict heavy FMC violation, because inserted foreign roots routinely pick up native bound morphemes in speech. The alternative is that Kazakh–Russian, in a written register, keeps its insertions bare, leaving the FMC mostly intact. These cannot both be right, and no one has measured it on this pair. What is missing is a quantitative test on Kazakh–Russian specifically, on written data, that codes the structure of switches, reports how often the classic constraints are respected, and handles the borrowing problem explicitly rather than ignoring it. That is the gap this study fills.

3. Data

The corpus is built from KazSAnDRA (IS2AI), an openly-licensed (CC BY 4.0) collection of 172,015 unique Kazakh online reviews. The register is the kind of short, evaluative, public writing in which Kazakh–Russian mixing actually occurs at high density: product, service, and media reviews written by ordinary users. An openly-licensed source was chosen deliberately so that the corpus, the code, and every coding decision can be released and rebuilt by anyone, which is rare in studies of this pair and is itself part of the contribution.

A sentence was treated as a candidate Kazakh–Russian code-switch when it satisfied three conditions: it contained at least two tokens with Kazakh-unique graphemes (so the matrix is genuinely Kazakh, not Russian), at least one Russian-marked token (a Russian-only grapheme such as ё/э/ъ, or unambiguous Russian inflectional morphology), and a Kazakh–Russian

adjacency at the switch site. This yielded 2,447 candidate sentences, about 1.4% of the source, a realistic rate for naturally occurring intra-sentential switching in written review text.

The candidate set was then cleaned in stages, because raw social-media text carries material that would distort a structural count. Non-lexical noise (keysmash, laughter, repeated characters) was removed, as were proper names (which are language-neutral and not switches in the relevant sense), and established borrowings were excluded by the diffusion criterion described in Section 4. Table 1 reports each step. After cleaning, 1,618 single-item lexical switches across 1,451 sentences remained as the set on which the Free-Morpheme Constraint was tested; multi-word spans were set aside for separate analysis.

Provenance and reproducibility are treated as first-class. Build scripts and a per-row provenance log are released, and the corpus is rebuildable from the public source. On ethics, the data are public review text; no handles or personal identifiers were collected; the work is minimal-risk and license-compliant. The author declares no conflict of interest.

4. Method

The analysis has four coding decisions, each fixed before any result was computed so that the study is a genuine test rather than a search for a favourable number.

Switch-structure coding. Each switch span was coded as either a single-word insertion or a multi-word span (an alternation or an embedded-language island). This distinction establishes whether the corpus is dominated by insertion (single foreign words dropped into a native frame) or by alternation (the grammar itself switching for a stretch), and it determines which constraints are even testable on a given item.

Borrowing exclusion. Because an established loanword carrying native morphology is not a constraint violation, a Russian-script item was treated as an established borrowing, and excluded from the FMC test, if its stem, after exhaustive Kazakh-suffix peeling, occurred at least five times in a 6.6-million-token monolingual-Kazakh reference corpus built from fineweb-2 Kazakh. The intuition is simple: a Russian-origin form that already appears frequently in Kazakh-only text has diffused into Kazakh and is no longer a momentary switch. The frequency threshold was swept (3 / 5 / 10 / 20) to confirm that the headline result does not depend on the exact cut-off. This operationalises the nonce-borrowing problem that Muysken [7] and Odagiri [3] both flag, and makes the boundary decision reproducible rather than intuitive.

FMC coding. A genuine (non-borrowing) Russian root carrying an unambiguously Kazakh bound morpheme (a case, number, or possessive suffix whose form does not coincide with any Russian inflection) was coded as an FMC violation. A bare Russian free form was coded as FMC respected. Endings that are ambiguous because they coincide with a Russian inflection (for example *-ы*) were not automatically counted as violations, which makes the violation estimate conservative: the procedure errs toward under-counting violations, not inflating them.

MLF and EC. The matrix language was identified by which language supplies the inflectional frame of the clause. The Equivalence Constraint is not testable for a single inserted content word, so single-word insertions were coded N/A for the EC; multi-word alternations were flagged for case-by-case coding.

The coding scheme and the analysis were pre-registered, written down before results existed, so the study reports honest research rather than a post-hoc fit. Proportions are reported with 95% Wilson confidence intervals, which behave well for the small violation count.

Coding validity. The automated coding was checked against a 50-item manual gold subsample coded by the author, a native Kazakh–Russian bilingual. Automated-versus-manual agreement on the FMC break/non-break decision was 98%. A second independent native coder and a formal Cohen's κ on a 15–20% subsample are identified as the planned reliability extension; they would strengthen the reliability claim and are noted as the next step before final publication. The single-coder validation is reported transparently rather than presented as full inter-rater reliability.

5. Results

Insertional dominance. Of 2,759 switch spans, 2,709 (about 98%) are single words; only 50 are multi-word. Kazakh–Russian online switching is therefore insertional rather than alternational: speakers drop individual Russian words into a Kazakh sentence far more often than they switch the grammar for a stretch.

Matrix language. Kazakh is the matrix language in 98.3% of sentences (95% CI 97.6–98.8). Kazakh reliably supplies the inflectional frame, and Russian supplies content. This is the stable-matrix profile the MLF model predicts and that the spoken neighbour studies also report.

Free-Morpheme Constraint. On the cleaned set of 1,618 single-item lexical switches the FMC is violated in 4.9% of cases (95% CI 3.9–6.0) and respected in 95.1%. The rate is stable across borrowing thresholds (4.7–5.1%) and barely moves between the raw and cleaned candidate sets, so it is not an artefact of where the borrowing line is drawn. The Free-Morpheme Constraint therefore largely holds for this pair in this register: Russian items are overwhelmingly inserted bare, without a Kazakh suffix attached to a switched Russian root. The violations that do occur cluster in slang and recent loans that take Kazakh morphology, for example *мэмбер-лер-ге* 'to the members' (from English *member* via Russian) and a vulgar slang neologism taking the Kazakh plural suffix *-мар*.

Coding validity. On the 50-item manual gold subsample, automated-versus-manual agreement on the FMC break/non-break decision was 98% (FMC-break precision 2/3, recall 2/2). The automated procedure tracks the human decision closely on this subsample, while the formal second-coder reliability test remains the planned extension.

Table 1. Corpus construction and the cleaning pipeline. Each row is a filtering step from the raw extraction to the set on which the Free-Morpheme Constraint was tested.

Step	Sentences	Switch tokens	Note
Code-switch candidates extracted from KazSAnDRA (172,015 reviews)	2,447	2,759 spans	2,709 single-word, 50 multi-word
First-pass borrowing exclusion (diffusion threshold ≥5)	2,006	2,202 single-word	507 first-pass borrowing tokens removed
Final clean set, exhaustive re-peel (single-item lexical switches)	1,451	1,618	−147 noise, −239 proper names, −544 borrowings

Note. The first-pass borrowing filter removes 507 tokens; the final recompute re-derives borrowings with exhaustive Kazakh-suffix peeling and removes 544, the authoritative count

behind the 1,618-item test set. The intermediate token totals therefore differ slightly between the two independent passes, and the 1,618-item set is the one on which all reported FMC rates are computed.

Table 2. Free-Morpheme Constraint on 1,618 single-item lexical switches. Proportions with 95% Wilson confidence intervals; the violation rate is stable across borrowing thresholds (4.7–5.1%).

Outcome	n	Rate	95% CI
FMC respected (bare insertion)	1,539	95.1%	94.0–96.1
FMC violated (Kazakh bound morpheme on Russian root)	79	4.9%	3.9–6.0

6. Discussion

6.1 The prediction and what the data show

The typological prediction, drawn from spoken agglutinative-pair studies, was that the Free-Morpheme Constraint would be the first constraint to fail for Kazakh–Russian. In Turkish–English, Kemaloğlu-Er [1] found that the most salient switching pattern was Turkish case, plural, and possessive suffixes attached to English noun stems, exactly the configuration the FMC forbids. In Arabic–English, Bahri [2] measured FMC violation at 34.5% of utterances, again through bound morphemes of the agglutinating language attaching to free morphemes of the other. If Kazakh–Russian behaved like these pairs, a violation rate in the same range would be expected.

The written Kazakh–Russian data do not behave that way. The FMC is violated in only 4.9% of single-item lexical switches, an order of magnitude below the Arabic–English spoken rate, and the result is stable across borrowing thresholds and across the raw and cleaned candidate sets. Speakers overwhelmingly insert Russian content words as bare free forms inside a Kazakh frame, which respects the FMC; attaching a Kazakh suffix to a switched Russian root, the configuration the FMC rules out, stays rare. The data agree with the closest neighbours on the two coarse facts (switching is insertional, and Kazakh is the matrix) but part company on the one fact the FMC turns on.

6.2 Why the contrast is register, not typology

The contrast is sharpest against Kyrgyz–Russian, the Turkic sister pair. Odagiri [3] found that the single most frequent insertion pattern in spoken Bishkek Kyrgyz–Russian was a Russian noun *with* a Kyrgyz suffix attached, the very configuration the written Kazakh data avoid about 95% of the time. Kazakh and Kyrgyz are about as alike as two languages get, so the contrast cannot be typological: if a shared agglutinative morphology drove morphological integration, both should show it. The most economical explanation is register and modality. The data here are written product and service reviews, a register of short evaluative sentences built around single noun and adjective insertions, where there is little morphological pressure to fold a Russian stem into a Kazakh case frame. The high-integration studies draw on spoken data: classroom interaction in Kemaloğlu-Er [1], conversational podcasts in Bahri [2], and recorded conversation in Odagiri [3]. In speech, longer and syntactically denser utterances create far more sites at which an agglutinating suffix must land on something, and sometimes that something is

a switched root. On this reading, the low Kazakh–Russian violation rate is a property of written review text, not of the language pair, and the quantitative result is best read as the first measurement of how written Kazakh–Russian behaves, to be set beside the spoken neighbours rather than merged with them.

6.3 Where the violations fall

A second observation reinforces the same conclusion: where the violations actually fall. The small set of genuine FMC violations is not random. It is concentrated in slang and very recent loans that have not yet stabilised as borrowings: *мэмбер-лер-ге* (from English *member* via Russian) and a vulgar slang neologism inflected with the Kazakh plural *-мар*. These are precisely the open-class, newly-entering items that Forker's [9] borrowing/code-switching part-of-speech hierarchy predicts will be integrated first, which suggests that morphological integration in Kazakh–Russian is an early stage of borrowing rather than a productive switching strategy. Both the register contrast and the lexical profile of the violations point the same way: in this written register, Kazakh–Russian contact is insertional and constraint-respecting, with morphological integration a minority, lexically-restricted pattern.

6.4 Fit with the Matrix Language Frame model

The finding fits an insertional MLF account, with Kazakh as a stable matrix supplying the inflectional frame and Russian contributing content words. It does not, on its own, settle whether the same would hold in speech: the modality reading above predicts it might not, and a parallel spoken Kazakh–Russian corpus is the natural next test. The result also has a practical implication for anyone building tools (spell-checkers, language-identification systems, or text normalisers) for Kazakh social media: in this register, Russian-origin material is far more likely to appear as a bare inserted word than as a morphologically integrated form, which is the easier case for a Kazakh-matrix parser to handle.

7. Limitations

The study is bounded in ways that should be read alongside the result. It covers a single register and a single source: online reviews. Spoken Kazakh–Russian may pattern closer to the Turkish–English and Arabic–English findings, and this study cannot rule that out. The candidate set carries the usual social-media noise (proper names, typos, keysmash). Although a cleaning pass removed 147 noise tokens, 239 proper names, and 544 borrowings, and the FMC rate barely changed (4.8% raw to 4.9% cleaned), some residue such as lowercase proper names and rare misspelled borrowings will remain until a full native pass. The borrowing/code-switch boundary is criterion-dependent: the threshold sweep bounds the sensitivity but does not eliminate the underlying continuum [7]. Reliability rests on a single coder (the author) with automated assistance and a 50-item validation; a second independent native coder and a formal Cohen's κ would strengthen the claim and are the planned next step before final publication. Finally, the Equivalence Constraint was not quantified for single insertions, and the multi-word subset awaits case-by-case coding.

8. Conclusion

In Kazakh–Russian online code-switching, contact is dominated by single-word lexical insertion into a Kazakh matrix, and the classic Free-Morpheme Constraint largely holds, with violation at only 4.9%. This runs against the prediction (drawn from spoken Turkish–English, Arabic–English, and even the sister pair Kyrgyz–Russian) that an agglutinative–fusional pair would break that constraint first through morphological integration. The result shares the insertional, stable-matrix profile documented for Kyrgyz–Russian, Sanzhi–Russian, and Ahıska

Turkish–Russian, but diverges from those spoken corpora precisely on morphological integration, which the written data keep rare. The most economical reading of that divergence is a register-and-modality effect rather than a typological one. The contribution is a first quantitative, reproducible test of these classic constraints on this under-studied pair, and the first measurement of how the pair behaves in writing. The clearest next step is a parallel spoken Kazakh–Russian corpus, coded the same way, to test whether the low violation rate is truly a property of the written register or of the pair itself.

References:

[1] Kemaloğlu-Er E. Patterns of intrasentential code-switching in Turkish–English bilingual discourse: Testing the Free Morpheme and the Equivalence Constraint // Artıbilim: Adana Science and Technology University Journal of Social Science. 2018. Vol. 1, No. 2. P. 35–45.

[2] Bahri S. Challenging the Universality of Poplack's Syntactic Constraints: Insights from Arabic–English Code-Switching among Saudi Arabian Women: Doctoral dissertation. Albuquerque: University of New Mexico, 2025. URL: https://digitalrepository.unm.edu/ling_etds/89/ (accessed: 24.06.2026).

[3] Odagiri N. Patterns and features of Kyrgyz–Russian code-switching in Bishkek // Journal of Foreign Language Studies (Kansai University). 2020. Vol. 23. P. 21–41.

[4] Poplack S. Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching // Linguistics. 1980. Vol. 18, No. 7–8. P. 581–618. DOI: 10.1515/ling.1980.18.7-8.581.

[5] Sankoff D., Poplack S. A formal grammar for code-switching // Papers in Linguistics. 1981. Vol. 14, No. 1. P. 3–45. DOI: 10.1080/08351818109370523.

[6] Myers-Scotton C. Duelling Languages: Grammatical Structure in Codeswitching. Oxford: Oxford University Press, 1993.

[7] Muysken P. Bilingual Speech: A Typology of Code-Mixing. Cambridge: Cambridge University Press, 2000.

[8] İnan K. Code-switching pattern of Turkish–Russian bilingual Ahıska Turks and the Matrix Language Frame model // Bilig. 2022. No. 103. P. 183–209. DOI: 10.12995/bilig.10307.

[9] Forker D. Sanzhi–Russian code switching and the Matrix Language Frame model // International Journal of Bilingualism. 2019. Vol. 23, No. 6. P. 1448–1468. DOI: 10.1177/1367006918798971.

Appendices

Appendix A. Data and reproducibility

Source series, build scripts (build_corpus_kazsandra.py, build_kz_lexicon.py, apply_borrowing_filter.py, constraint_code_v3.py, clean_and_recompute.py), the provenance JSON, and the analysis outputs are released for rebuild from the public KazSAnDRA source. A reader can reconstruct the candidate set, the borrowing filter, and the final FMC counts from the released package.

Appendix B. Coding manual and validation

The codebook (operational definitions of single-word insertion, established borrowing, FMC violation, and matrix language, with decision rules), the 200-sentence native coding sheet, and the 50-item author validation gold (98% break/non-break agreement) are included in the replication package.

UDC 551.583:551.46

Demin Mark

Grade 9 student, Nazarbayev Intellectual School
(Atyrau, Kazakhstan)

IS THE CASPIAN SEA WARMING FASTER THAN THE WORLD OCEAN, AND IS THE TREND ACCELERATING? A SATELLITE-ERA ANALYSIS OF SEA-SURFACE TEMPERATURE (1982–2022)

Abstract: The Caspian Sea is the largest enclosed body of water on Earth and borders Kazakhstan, yet most reporting on ocean warming uses the global average and says little about how fast individual enclosed basins are warming. Enclosed seas are shallow and hold a small volume of water for their surface area, so they are expected to respond quickly to a warming atmosphere, but for any specific basin this should be tested against the global rate rather than assumed. This study uses open satellite-era sea-surface-temperature (SST) data from the NOAA Optimum Interpolation analysis to test two questions over 1982–2022: whether the Caspian Sea warmed faster than the global ocean, and whether its warming accelerated. The Caspian was isolated from surrounding land using the dataset's land–sea mask (57 water cells), the seasonal cycle was removed against a fixed 1991–2020 baseline, and a linear warming rate was estimated with autocorrelation-corrected standard errors and a non-parametric Theil–Sen check. Over 1982–2022 the Caspian warmed at +0.24 °C per decade (95% CI 0.16 to 0.32), about 2.2 times the global-ocean rate of +0.11 °C per decade (95% CI 0.09 to 0.13); the difference was statistically significant (+0.13 °C per decade, 95% CI 0.04 to 0.21). The global rate matched published satellite-era estimates, which validates the method. The Caspian's warming did not accelerate within the period: the quadratic time term was not significant, and the rate was lower in the second half (+0.09 °C per decade) than the first (+0.43 °C per decade). The Caspian therefore warmed markedly faster than the world ocean over the satellite era but did so without accelerating. Because the analysis is observational, it describes the trend and its uncertainty rather than attributing it to a single cause.

Keywords: Caspian Sea, sea-surface temperature, climate change, warming trend, satellite data, OISST, enclosed sea

Introduction. The world ocean has warmed measurably over the satellite era. Across widely used sea-surface-temperature (SST) datasets, the global ocean between 60° S and 60° N warmed by roughly 0.11 to 0.18 °C per decade over the satellite period, and the average surface temperature of the ocean has risen by close to 0.9 °C since the late nineteenth century [1, 2]. These figures are almost always reported as a global or near-global average. They say little about how fast any one regional sea is warming, and regional rates can depart substantially from the global mean.

Enclosed and semi-enclosed seas are a case where the regional rate is expected to differ. Such basins are shallow and hold a small volume of water for their surface area, so the same heat input from the atmosphere is spread through less water and is not easily exported to the open ocean. On physical grounds this should make an enclosed sea respond faster to a warming

atmosphere than the global ocean does. The expectation is intuitive, but for any specific basin it should be checked against data rather than assumed.

The Caspian Sea is the largest enclosed body of water on Earth and is directly relevant to Kazakhstan, which forms much of its northern and eastern shore [3]. The northern Caspian in particular is very shallow, which sharpens the physical reason to expect rapid warming there. The Caspian has been studied intensively for its falling water level and for its wind and wave climate, and reanalysis and satellite products have been validated for use over the basin in those studies [4, 5]. Its surface-temperature trend, placed side by side with the global ocean rate and tested for acceleration, is a narrower and more checkable question, and it is the one this study addresses.

Two specific questions are asked. First, did the Caspian Sea warm faster than the global ocean over the satellite era? Second, did the Caspian's warming accelerate within that period, rather than proceeding at a steady rate? Both questions are answered with open data and a transparent method, and both are framed so that the answer can be a clear yes, no, or only-under-certain-conditions rather than an open-ended description.

Materials and Methods. Dataset. Monthly SST was taken from the NOAA Optimum Interpolation SST analysis, version 2 (OISST v2), a 1° global product that blends satellite and in-situ observations and is distributed by the NOAA Physical Sciences Laboratory [6, 7]. The version used here covers December 1981 to January 2023; the analysis was restricted to the complete calendar years 1982 through 2022 (492 monthly fields). OISST v2 was chosen because it provides continuous monthly coverage over enclosed basins, including the Caspian, and because it is distributed together with a land–sea mask that allows water cells to be separated from land.

Isolating the Caspian. A bounding box alone is not sufficient, because OISST v2 reports filled values over land as well as water, so an unmasked box average would be contaminated by land. To avoid this, the Caspian was defined as the intersection of a bounding box (36–48° N, 46–55° E, a range that contains no other large water body) with the dataset's land–sea mask, keeping only the 57 grid cells flagged as water. The land–sea mask was verified directly: known land points (interior Kazakhstan, the Turkmen desert, the Zagros) were correctly flagged as land, and Caspian interior points as water. Regional means were area-weighted by the cosine of latitude.

Global comparison region. The global-ocean baseline was the area-weighted mean of all water cells between 60° S and 60° N, the latitude band conventionally used to avoid sea-ice contamination of the trend [1]. The Caspian cells are included in this global mean but make up 0.18 % of all ocean cells, so their contribution to the global value is negligible.

Pre-processing. For each region the monthly field was averaged to a single monthly series, and the seasonal cycle was removed by subtracting the long-term mean of each calendar month, computed over a fixed 1991–2020 climatological baseline applied identically to both regions. Using the same baseline period for the Caspian and the global ocean is essential so that the two trends are comparable.

Trend estimation. The warming rate for each region was estimated as the slope of an ordinary-least-squares regression of the monthly anomaly on time, expressed in °C per decade. Because monthly SST anomalies are autocorrelated, the standard error of the slope was corrected for serial correlation with Newey–West (HAC) standard errors using twelve lags,

which widens the confidence interval to a realistic width. The non-parametric Theil–Sen slope with a Mann–Kendall test was computed as a robustness check that is insensitive to outliers.

Acceleration test. Acceleration was tested two ways. First, a quadratic term in time was added to the regression; a significant positive quadratic coefficient would indicate an accelerating trend. Second, the record was split at its midpoint and the linear rate of the first half (1982–2002) was compared with that of the second half (2002–2022), with a confidence interval on the difference. Reporting both guards against reading a single curved fit as acceleration when it is not robust.

Regional comparison. The central comparison is the difference between the Caspian decadal rate and the global decadal rate, reported with a confidence interval, so that the answer to the first research question is not just which number is larger but whether the difference is distinguishable from zero given the noise in the data.

Reproducibility and ethics. The analysis uses only publicly available, de-identified environmental data and involves no human or animal subjects, so no ethical approval was required. The dataset versions, access date, and full analysis code are preserved, so every number below can be reproduced from the two open source files.

Results. Caspian and global warming rates. Over 1982–2022 the Caspian Sea warmed at +0.239 °C per decade, more than twice the global-ocean rate of +0.109 °C per decade (Table 1). Both trends were highly significant, and both were confirmed by the non-parametric Theil–Sen estimator (+0.247 and +0.110 °C per decade respectively). The global rate falls within the published satellite-era range of 0.11–0.18 °C per decade [1], which validates the processing pipeline against an independent benchmark.

Table 1 – Linear sea-surface-temperature trends for the Caspian Sea and the global ocean over 1982–2022, with autocorrelation-corrected (Newey–West) confidence intervals.

Region (OISST v2, 1982–2022)	Warming rate, °C/decade	95% CI	p-value
Caspian Sea	+0.239	+0.155 to +0.323	< 0.001
Global ocean (60° S–60° N)	+0.109	+0.090 to +0.128	< 0.001
Difference (Caspian − global)	+0.129	+0.043 to +0.215	< 0.01

Direct comparison. The difference between the two rates was +0.129 °C per decade, and its 95 % confidence interval (0.043 to 0.215) excludes zero. The Caspian Sea therefore warmed significantly faster than the global ocean over the satellite era, by roughly a factor of 2.2. This answers the first research question: yes.

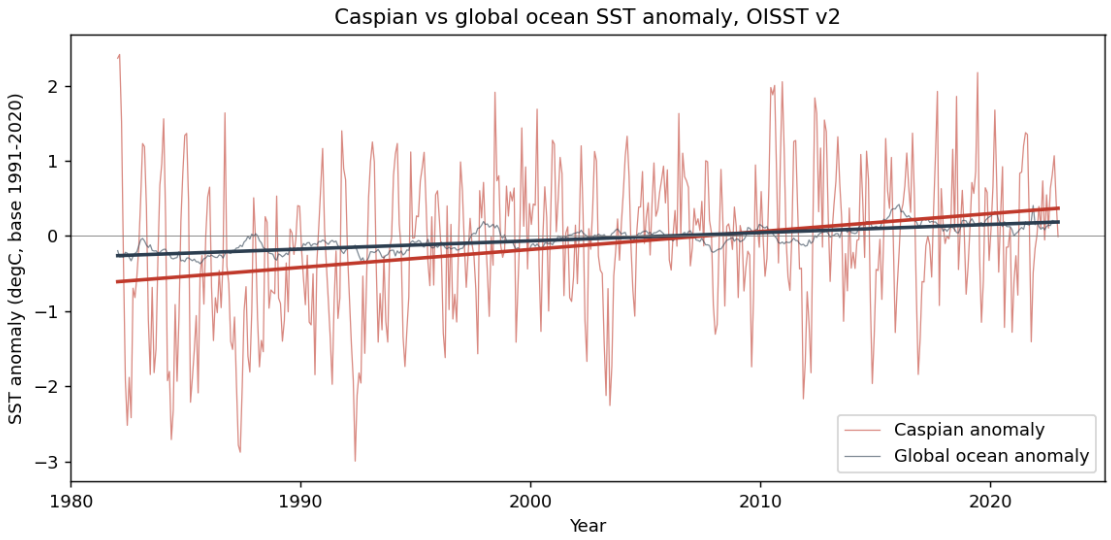


Figure 1 – Monthly SST anomaly for the Caspian Sea and the global ocean (1982–2022), relative to the 1991–2020 baseline, with fitted linear trends. The Caspian series is both warmer-trending and far noisier than the global mean.

Acceleration. The Caspian's warming did not accelerate (Table 2). The quadratic time term was small, negative, and not significant ($p = 0.095$), so there is no statistical support for an upward curve. The split-half test pointed the other way: the warming rate was higher in the first half of the record (+0.432 °C per decade, 1982–2002) than in the second (+0.088 °C per decade, 2002–2022), a decrease of −0.345 °C per decade whose confidence interval (−0.679 to −0.010) just excludes zero. The two tests agree that there was no acceleration; if anything the rate eased in the second half, though this easing is borderline. This answers the second research question: no.

Table 2 – Tests for acceleration of the Caspian warming trend. Neither test supports acceleration.

Caspian acceleration test	Estimate	95% CI	p-value
Quadratic time coefficient (°C/yr²)	−0.00068	—	0.095
First-half rate (1982–2002), °C/decade	+0.432	—	—
Second-half rate (2002–2022), °C/decade	+0.088	—	—
Change between halves, °C/decade	−0.345	−0.679 to −0.010	< 0.05

Discussion. Over 1982–2022 the Caspian Sea warmed about 2.2 times as fast as the global ocean, and the difference was statistically significant. This supports the physical expectation that a large but shallow enclosed sea, with a small water volume for its surface area, warms faster than the open ocean, and it attaches a specific number to that expectation for a basin on Kazakhstan's coast. The result is strengthened by the fact that the same pipeline

reproduced the known global rate of about 0.11 °C per decade, so the Caspian figure is not an artefact of the processing.

The second finding is the more surprising one. Despite warming quickly overall, the Caspian's warming did not accelerate within the period; the rate was actually lower in the second half of the record than the first. This runs against a simple picture in which a warming world drives an ever-faster regional trend, and it suggests that the Caspian's surface temperature is shaped by more than the global background, plausibly by its own variability in level, inflow, and circulation, which are known to be large for this basin [4, 5]. The present data establish that the warming did not speed up; they do not by themselves explain why, and that limit is stated rather than filled with speculation.

This study has clear limitations. First, it uses a single SST product. The global sanity check gives confidence in the pipeline, but a stronger version of this analysis would repeat it with an independent product to confirm that the Caspian rate and the no-acceleration result are not specific to OISST; recent work shows that satellite-era SST trends can differ in consequential ways across datasets [1]. Choosing that second product is not trivial for an enclosed basin: two widely used open datasets were examined and found unsuitable here, because a coarse open-ocean reconstruction does not resolve the basin well and a second gridded product was found to carry only a repeating climatology over the Caspian rather than real year-to-year values. A meaningful cross-validation therefore requires a product that genuinely represents the enclosed sea, such as the ERA5 reanalysis or a Copernicus Marine analysis, and running it is the clearest next step. Second, the analysis covers the surface temperature only, not heat content at depth, water level, or salinity, all of which matter for the Caspian. Third, the observing network is sparser over an enclosed sea than over the open ocean, so the regional values carry more uncertainty than a global average. Finally, the design is observational: it characterizes the trend and its uncertainty but does not attribute the warming to a single driver, and the result is specific to the Caspian and to 1982–2022.

Conclusion. Using open satellite-era data and a transparent, pre-specified method, this study finds that the Caspian Sea warmed at about +0.24 °C per decade over 1982–2022, roughly 2.2 times the global-ocean rate, a difference that is statistically significant. It also finds that this rapid warming did not accelerate within the period and, if anything, eased in the second half. The Caspian is therefore warming faster than the world ocean but not at an increasing rate. The method is fully reproducible from open data and can be reapplied to other enclosed seas or extended as longer records and additional SST products are brought in.

References:

1. Yamaguchi R., et al. Consequential differences in satellite-era sea surface temperature trends across datasets. Nature Climate Change, 2025.

2. Copernicus Climate Change Service. Sea surface temperature — climate indicator. ECMWF / C3S, 2024.

3. Kostianoy A.G., Kosarev A.N. (eds.). The Caspian Sea Environment. The Handbook of Environmental Chemistry, Vol. 5P. Berlin: Springer; 2005.

4. Chen J.L., Pekker T., Wilson C.R., Tapley B.D., Kostianoy A.G., Crétaux J.-F., Safarov E.S. Long-term Caspian Sea level change. Geophysical Research Letters, 2017;44(13):6993–7001. doi:10.1002/2017GL073958.

5. Wind and wave climate of the Caspian Sea from ERA5-driven hindcast, 1982–2023. ScienceDirect, 2024.

6. Reynolds R. W., et al. Daily high-resolution-blended analyses for sea surface temperature. Journal of Climate, 2007;20:5473–5496.

7. Huang B., et al. Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) version 2.1. Journal of Climate, 2021;34:2923–2939.

УДК 004.85:81'322

Assan Ramazan
11th-grade student
Nazarbayev Intellectual school
(Shymkent, Kazakhstan)

KAZSWITCH: A CONTROLLED BENCHMARK FOR KAZAKH–RUSSIAN CODE-SWITCHING IN MULTILINGUAL LANGUAGE MODELS

Abstract: Large language models are usually tested on clean, single-language text, but hundreds of millions of bilingual speakers mix two languages inside a single sentence, a behaviour called code-switching. In Kazakhstan, Kazakh and Russian are routinely combined this way, yet almost no benchmark measures how well models handle Kazakh–Russian code-switched input, so the size of any performance loss is unknown. This study introduces KazSwitch, a controlled benchmark that isolates the effect of code-switching itself. Each test item is written three ways with identical meaning and an identical correct answer: pure Kazakh, pure Russian, and Kazakh–Russian code-switched. Because only the mixing changes across the three versions, any drop in a model's accuracy on the code-switched version is attributable to code-switching rather than to topic or difficulty. The pilot benchmark comprises 26 base items in each of the three versions: 10 authored and verified by a native Kazakh–Russian bilingual, and 16 derived from a published Kazakh–Russian code-switching corpus (reused verbatim under CC BY-SA 4.0) by adding a factual numeric question whose answer is identical across the three versions. A fixed panel of three open compact multilingual models (2–4B parameters: gemma2:2b, llama3.2:3b, qwen2.5:3b) was evaluated locally at temperature zero, and performance was summarized by the Code-Switching Penalty (CSP), defined as monolingual accuracy minus code-switched accuracy. Pooled across the three models with an item-clustered bootstrap, the penalty is +0.167 (95% CI +0.045 to +0.295), a statistically significant drop of about seventeen accuracy points caused by the mixing alone; errors were dominated by mis-resolving the switched key term rather than by wrong-language or refusal responses. KazSwitch is released as an open, versioned dataset and reproducible code so that the comparison can be rerun for free and extended to a larger, direction-balanced set and to other language pairs.

Keywords: code-switching, language models, Kazakh, Russian, benchmark, multilingual NLP, evaluation

Introduction. Code-switching, the use of two languages inside a single conversation or even a single sentence, is one of the most common ways bilingual people actually speak. It is the everyday register for hundreds of millions of speakers across language pairs such as Spanish–English, Hindi–English, and Arabic–French, and it is the norm rather than the exception in much of the world. Large language models, however, are trained and evaluated mostly on clean, single-language text. When their abilities are measured, the test sets are usually monolingual, so a model can look strong on standard benchmarks while still struggling on the mixed-language input that bilingual users produce naturally.

In Kazakhstan, Kazakh and Russian are mixed constantly, both in speech and in writing online. This makes Kazakh–Russian a natural and important case for studying code-switching, and it is under-served: while there are now Kazakh and Russian language resources, very little exists that measures how models handle the two languages mixed together in a controlled way.

The practical consequence is that no one can currently say how much accuracy a model loses on Kazakh–Russian code-switched input, because there is no benchmark designed to measure exactly that.

Existing code-switching benchmarks established the field but do not fill this gap. LinCE assembled corpora for four language pairs across tasks such as language identification, named-entity recognition, and sentiment analysis [1], and GLUECoS did the same for English–Hindi and English–Spanish across a similar task suite [2]. More recent work targets large language models directly: CodeMixBench evaluates code-mixing across eighteen languages and several reasoning tasks [3], and other 2025 studies probe how well LLMs understand code-switched text [4]. None of these covers Kazakh–Russian, and, more importantly, none isolates the effect of switching with a controlled design in which the same item is presented monolingually and code-switched.

For Kazakh–Russian specifically, the nearest existing resources address different problems. A 2025 study released a Kazakh–Russian code-switching parallel corpus, but it is built for machine translation and is evaluated with translation metrics rather than as a controlled comprehension benchmark [5]. The Qorgau benchmark evaluates the safety of models in Kazakh and Russian, including code-switched prompts, but its target is safety behaviour, not general accuracy [6]. KazMMLU measures knowledge in Kazakh and Russian, but each item is monolingual and the benchmark is not concerned with switching [7]. KazSwitch occupies the slot none of these fills: a controlled benchmark that measures how much accuracy is lost specifically because the input is code-switched.

The study asks two questions. First, how large is the Code-Switching Penalty — the drop in accuracy from monolingual to code-switched input — across a small panel of open compact multilingual models? Second, what are the most common, repeatable ways these models fail on code-switched input? A third question of interest, whether the penalty depends on switch direction, is defined but deferred, because the pilot item pool is almost entirely Kazakh-matrix and cannot answer it. The contribution is twofold and stated precisely: a native Kazakh–Russian bilingual author constructed the 10 native controlled triples and designed and ran the whole evaluation, and the benchmark is then expanded with 16 further items derived from a published Kazakh–Russian code-switching corpus (reused under licence and attributed, not presented as native-authored) to give the pooled estimate more power. Together they yield the first controlled measurement of this penalty for Kazakh–Russian, built so that the same protocol can be reused for a larger, direction-balanced set and for any other language pair.

Materials and Methods.

Design: controlled minimal triples. The core of KazSwitch is that every test item exists in three versions with identical meaning and an identical correct answer: pure Kazakh, pure Russian, and Kazakh–Russian code-switched. Holding meaning, topic, and answer constant across the three versions means that the only thing changing is the language mixing, so a difference in a model's accuracy across versions can be attributed to code-switching itself and not to a harder topic or a different question. This is the methodological backbone that makes KazSwitch a controlled study rather than a leaderboard.

Task format. The task is short-answer comprehension: each item is a short passage and a question with a single, unambiguous answer that is graded by normalized exact match (case- and whitespace-insensitive, punctuation removed, with a small accepted-variants list per item). A response is scored correct when the normalized gold answer (or an accepted variant) appears

as a complete token in the normalized response; whole-token rather than raw-substring matching is used so that a short gold answer such as "5" or "бес" (five) is not spuriously credited inside a longer wrong response such as "50" or "бесконечность" (infinity). Automatic exact-match grading is used because it removes grader subjectivity and makes the comparison across the three versions clean.

Item construction. The code-switched material is real human usage, not synthesized; generating the code-switched items with a language model was deliberately rejected, because the test would then measure how models handle model-produced switching rather than genuine human switching. The 26-item pilot is built from two honest sources, recorded per item. Ten items were authored and native-verified by a Kazakh–Russian bilingual: for each, the author wrote the three meaning-equivalent versions and a single short comprehension question with one correct answer, and recorded the switch direction. The other sixteen items are derived from an existing Kazakh–Russian code-switching corpus [5]: the three sentence versions (code-switched original, pure Kazakh, pure Russian) are reused verbatim from that corpus, which was produced by its own native annotators, under its CC BY-SA 4.0 licence, and the only thing added is a factual numeric question whose single correct answer appears identically in all three versions. These corpus-derived items are not presented as native-authored; they are attributed to the source corpus, and because the answer is a number present identically in all three versions, no language-specific judgement is required to confirm meaning-equivalence for them. The dataset is fixed and versioned: 26 base items × 3 versions = 78 instances.

Models. A small, fixed panel of three open compact multilingual models is evaluated: gemma2:2b, llama3.2:3b, and qwen2.5:3b (2–4 billion parameters), run locally through an OpenAI-compatible endpoint. The panel is deliberately limited to small open-weight models so that the entire benchmark can be reproduced for free on consumer hardware, with no paid API and no dependence on a closed model whose version may change; this is a reproducibility choice, and it scopes the result to open compact models rather than to frontier systems. All models are run at temperature zero for determinism, fixed versions are recorded, and every prompt and raw response is released. No model is trained or fine-tuned for the task; the evaluation is zero-shot. Each model is shown only the passage and the single question, never the gold answer, the accepted variants, or the other-language versions of the item.

Primary metric and analysis. The headline metric is the Code-Switching Penalty (CSP), monolingual accuracy minus code-switched accuracy, computed within item so that it is a paired difference. Because the penalty for any single model is underpowered at a pilot of this size, the primary estimate is the CSP pooled across all three models, with a 95% confidence interval from an item-clustered bootstrap (items are resampled with all models retained, since the items are the authored unit and the main source of sampling variation); per-model CSPs are reported only as a secondary, descriptive breakdown with their own intervals and are not interpreted as a ranking. The pooled penalty is tested against the null of zero, with significance read from whether the bootstrap interval excludes zero.

Secondary analysis. Failure types are characterized with a coding scheme defined in advance (answered in the wrong language, copied a fragment instead of answering, mis-resolved a switched key term or otherwise wrong, or refused), applied to the incorrect code-switched responses, and the distribution is reported. Direction asymmetry was specified as a question of interest but is not tested in this pilot: the item pool is almost entirely Kazakh-matrix (25 of 26 items), so any Russian-matrix estimate would rest on a single item and would be noise. No

directional claim is made; a direction-balanced version is required to answer that question and is left to future work.

Analysis validity pre-flight. Before any model was run, three checks were fixed: the three versions of each item must share a single correct answer, so the comparison is valid (for the native-authored items this is confirmed by the bilingual author; for the corpus-derived items the answer is a number present identically in all three verbatim versions); accuracy must be compared within items (paired), not across independent samples; and the failure-type scheme must be defined in advance so that the taxonomy is not fitted to the outputs after the fact. The full protocol, including the metric and the failure-type scheme, was written before generating any results.

Ethics and disclosure. The benchmark is built from public, non-sensitive text and involves no personal or sensitive data. The code-switched sentences used in the corpus-derived items are reused verbatim from a published Kazakh–Russian code-switching corpus [5], cited and used under its CC BY-SA 4.0 licence; the native-authored items were written for this study. AI assistance is disclosed: language models were used to help draft manuscript prose, checked by the author, but the code-switched items — the measured variable — are real human usage and were not model-generated. The released dataset and code allow the evaluation to be reproduced independently on consumer hardware.

Reproducibility. The benchmark and evaluation are released so the result can be rerun for free without any paid service. The released material is: the item set with per-item provenance (each item tagged as native-authored or corpus-derived, with its source row), the complete set of 234 raw model responses, the evaluation script that computes accuracy and the item-clustered penalty, and the offline robustness script that reproduces the subtype and grading checks deterministically. Because the panel is open-weight and run locally at temperature zero, an independent reader can regenerate every number in this paper on consumer hardware.

Results. This section reports the output of the locked protocol above. Every value is computed from the run's 234 raw model responses, which are released in full; none is estimated.

Benchmark composition. The pilot benchmark comprises 26 base items, each in three versions (pure Kazakh, pure Russian, and Kazakh–Russian code-switched), for 78 instances per model and 234 model responses in total across the three models. Of the 26 items, 10 were authored and native-verified by a Kazakh–Russian bilingual and 16 were derived from an existing Kazakh–Russian code-switching corpus [5] by attaching a factual numeric question whose single answer appears identically in all three versions (the three sentence versions are reused verbatim from that corpus under its CC BY-SA 4.0 licence). Because the corpus from which the code-switched originals are drawn is predominantly Kazakh-matrix, the item pool is heavily skewed by switch direction (25 Kazakh-matrix, 1 Russian-matrix); this is a known limitation and is why no direction-asymmetry result is reported (see Limitations).

Code-Switching Penalty by model. Across the three open compact models, monolingual accuracy ranged from 0.635 to 0.827 and code-switched accuracy from 0.462 to 0.654, so every model was less accurate on the code-switched version of the same items than on the monolingual versions (Table 1). The pooled Code-Switching Penalty, estimated across all three models with an item-clustered bootstrap, is +0.167 (95% CI +0.045 to +0.295); because the interval excludes zero, the penalty is statistically significant. This is the study's headline result: on this controlled comparison, current open compact (2–4B) multilingual models lose on the order of seventeen

accuracy points specifically because the input mixes Kazakh and Russian rather than being in a single language.

Per-model penalties are reported as a secondary, descriptive breakdown and are individually underpowered at this pilot size; they should not be read as a ranking. llama3.2:3b showed the largest penalty (+0.231, 95% CI +0.019 to +0.423) and qwen2.5:3b a comparable one (+0.173, 95% CI +0.019 to +0.346), each with an interval excluding zero; gemma2:2b showed a smaller penalty (+0.096, 95% CI −0.058 to +0.250) whose interval includes zero, so for that single model the penalty is not distinguishable from zero at this N. The pooled estimate, not any single model, carries the claim.

Table 1 – Accuracy on monolingual versus Kazakh–Russian code-switched items, and the resulting Code-Switching Penalty, per model and pooled. The pooled penalty is the headline estimate; per-model rows are descriptive and underpowered at this pilot size.

Model	Monolingual accuracy	Code-switched accuracy	CSP (penalty)	95% CI
gemma2:2b	0.635	0.538	+0.096	−0.058 to +0.250
llama3.2:3b	0.692	0.462	+0.231	+0.019 to +0.423
qwen2.5:3b	0.827	0.654	+0.173	+0.019 to +0.346
Pooled (all models, item-clustered)	—	—	+0.167	+0.045 to +0.295

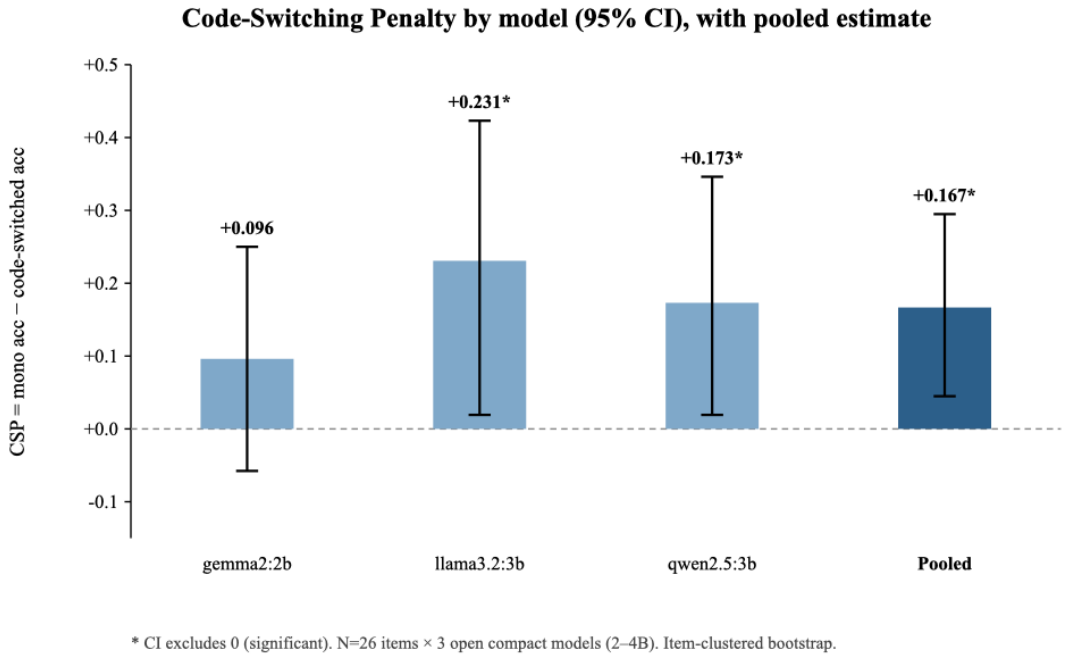


Figure 1 – Code-Switching Penalty by model and pooled, with 95% confidence intervals (item-clustered bootstrap). Bars whose interval excludes zero are statistically significant; the pooled estimate carries the claim.

Robustness 1: the penalty does not depend on the item source. Because the benchmark mixes 10 native-authored items with 16 corpus-derived items, the pooled penalty was recomputed separately within each source (Table 3). The penalty is positive in both — native items +0.233 (95% CI −0.017 to +0.450) and corpus-derived items +0.125 (95% CI −0.010 to +0.281) — so the direction of the effect is the same regardless of who wrote the item. Each subset, at 10 and 16 items, is individually underpowered and its interval includes zero; significance is reached only when the two are pooled, which is the expected pattern for a pilot at this size. The point is that the pooled result is not an artifact of one item source: both halves push the same way.

Table 3 – Code-Switching Penalty by item source. The effect is positive in both subsets; significance is reached only when pooled, as expected at this pilot size.

Item source	Items	CSP	95% CI	Significant
Native (Ramazan-authored)	10	+0.233	−0.017 to +0.450	no
Corpus-derived (KRCS)	16	+0.125	−0.010 to +0.281	no
All (pooled)	26	+0.167	+0.045 to +0.295	yes

Robustness 2: the conclusion does not depend on grading strictness. The headline uses strict, pre-specified exact-match grading. To test whether that strictness manufactured the result, every response was re-graded under a deliberately over-generous numeric rule that credits any answer containing the gold digit-string, even when the answer is genuinely wrong (this lenient rule wrongly accepts, for example, "300,000" for a gold answer of 300, "5*5 = 25" for 5, and a Chinese-script "5百万" hallucination for 5). Even under that over-credit, the pooled penalty is +0.160 (95% CI +0.045 to +0.276) and remains significant. The penalty therefore survives the most generous grading one could justify, so it is not an artifact of strict matching. The opposite bias is the one that matters here: strict exact-match slightly *over*-states the penalty, because it fails to credit correct code-switched answers in which the number carries a Kazakh case suffix (for example "5000тг", "100т", "15-и"); the true penalty therefore sits at the lower, still-significant end of the reported range. A suffix-aware numeric matcher is left as future work.

Failure taxonomy. Among incorrect code-switched responses, errors were classified with the pre-defined coding scheme. The great majority were cases where the model mis-resolved a switched key term or produced an otherwise wrong answer (0.83 of code-switched errors), and the remainder were cases where the model copied a fragment of the passage instead of answering the question (0.17); no errors in this pilot fell into the wrong-language or refusal categories (Table 2). The dominant failure mode is therefore comprehension breakdown at the switch, not a refusal or a language-selection error.

Table 2 – Distribution of failure types on incorrect code-switched items.

Failure type	Share of code-switched errors
Mis-resolved a switched key term / other wrong answer	0.83
Copied a fragment instead of answering	0.17
Answered in the wrong language	0.00
Refusal / no answer	0.00

Discussion. The central question was how large the Code-Switching Penalty is. The controlled comparison gives a clear answer for the models tested: pooled across the three open compact models, accuracy drops by about seventeen points (pooled CSP +0.167, 95% CI +0.045 to +0.295) when the same item is presented code-switched rather than in a single language. Because meaning, topic, and the correct answer are held constant across the three versions of each item, this drop is attributable to the language mixing itself and not to a harder question. The result puts a first measured number on a gap that, for Kazakh–Russian, could previously only be assumed.

The size of the effect should be read with the pilot's scope in mind. The claim rests on the pooled estimate; the per-model penalties are descriptive and individually underpowered, and one of the three models (gemma2:2b) had a penalty whose interval includes zero. The honest reading is therefore that a meaningful penalty is present in this class of model on average, not that any particular model's penalty has been pinned down. What makes the pooled penalty credible despite the small item count is the consistency of the direction across all three models: every model was less accurate on the mixed version. The two robustness checks reinforce this. Splitting by item source, the penalty is positive in both the native and the corpus-derived halves, so it is not produced by one part of the benchmark; and re-grading under a deliberately over-generous numeric rule leaves the pooled penalty significant, so it does not depend on strict matching. There is also visible heterogeneity between models that should not be over-read at this size: qwen2.5:3b, for instance, shows a large penalty on the native items but close to none on the corpus-derived ones. With ten and sixteen items per subset this is reported as a descriptive observation, not a finding, and it is one of the things a larger benchmark would resolve.

The failure taxonomy makes the score actionable. The errors did not concentrate in answering in the wrong language or in refusals; instead the dominant pattern (0.83 of code-switched errors) was mis-resolving the switched key term or otherwise producing a wrong answer, with a minority copying a passage fragment instead of answering (0.17). This points to comprehension breakdown at the switch as the mechanism, rather than a language-selection or safety-style failure, and it gives future work a concrete target: the models can read each language, but lose the thread when the answer-bearing material sits across a switch. Because the taxonomy was defined before the outputs were read, this distribution reflects the data rather than a post-hoc story.

The main limitations are size, item composition, and model scope. This is a 26-item pilot, so the absolute penalty is an estimate with a wide interval and the per-model figures are not a ranking. The item pool is strongly skewed by switch direction (25 Kazakh-matrix, 1 Russian-matrix), because the code-switched originals were drawn from a predominantly Kazakh-matrix corpus; for that reason this study makes no claim about direction asymmetry, and testing whether the penalty depends on switch direction is left to a future, direction-balanced version. The evaluated panel is deliberately limited to open compact (2–4B) multilingual models, chosen so the benchmark can be reproduced for free on consumer hardware; the penalty reported is therefore a statement about small open models and should not be extrapolated to frontier systems, which were not tested. The benchmark mixes 10 native-authored items with 16 corpus-derived numeric items, and it measures zero-shot behaviour of a fixed model set at one point in time; results may shift as models change. One grading limitation is disclosed because it cuts against the headline rather than for it: the strict exact-match normaliser does not credit a numeric answer fused to a trailing Kazakh case suffix (such as "5000тг" or "100т"), so it under-counts

some correct code-switched answers and therefore slightly inflates the measured penalty; the robustness re-grade shows the effect survives this, but a suffix-aware numeric matcher would tighten the estimate and is listed as future work. These limits are the cost of a controlled, reproducible pilot, and they are stated so the contribution is not overclaimed. The construction protocol transfers directly to a larger, direction-balanced set and to other language pairs, which is the main way the work can grow.

Conclusion. KazSwitch is a controlled benchmark that isolates and measures the accuracy cost of Kazakh–Russian code-switching for language models, using minimal triples in which only the mixing changes across the three versions of each item. In this pilot, three open compact multilingual models lost about seventeen accuracy points on code-switched input relative to the same items in a single language (pooled CSP +0.167, 95% CI +0.045 to +0.295), and their errors were dominated by mis-resolving the switched key term rather than by language-selection or refusal failures. The dataset and code are released openly so the measurement can be reproduced on consumer hardware, and the same protocol can be applied to a larger, direction-balanced set and to any other code-switched language pair.

References:

1. Aguilar G., Kar S., Solorio T. LinCE: A Centralized Benchmark for Linguistic Code-switching Evaluation // Proceedings of the 12th Language Resources and Evaluation Conference (LREC). — Marseille, 2020. — P. 1803–1813.

2. Khanuja S., Dandapat S., Srinivasan A., Sitaram S., Choudhury M. GLUECoS: An Evaluation Benchmark for Code-Switched NLP // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL). — 2020. — P. 3575–3585.

3. Yang Y., Chai Y. CodeMixBench: Evaluating Code-Mixing Capabilities of LLMs Across 18 Languages [Электронный ресурс]. — arXiv:2507.18791, 2025. — URL: https://arxiv.org/abs/2507.18791 (дата обращения: 26.06.2026).

4. Mohamed A., Zhang Y., Vazirgiannis M., Shang G. Lost in the Mix: Evaluating LLM Understanding of Code-Switched Text [Электронный ресурс]. — arXiv:2506.14012, 2025. — URL: https://arxiv.org/abs/2506.14012 (дата обращения: 26.06.2026).

5. Borisov M., Kozhirbayev Z., Malykh V. Low-resource Machine Translation for Code-switched Kazakh-Russian Language Pair [Электронный ресурс] // Proceedings of the NAACL 2025 Student Research Workshop. — 2025. — arXiv:2503.20007. — URL: https://arxiv.org/abs/2503.20007 (дата обращения: 26.06.2026).

6. Goloburda M., Laiyk N., Turmakhan D. [и др.] Qorgau: Evaluating LLM Safety in Kazakh-Russian Bilingual Contexts [Электронный ресурс] // Findings of the Association for Computational Linguistics: ACL 2025. — 2025. — arXiv:2502.13640. — URL: https://arxiv.org/abs/2502.13640 (дата обращения: 26.06.2026).

7. Togmanov M., Mukhituly N., Turmakhan D. [и др.] KazMMLU: Evaluating Language Models on Kazakh, Russian, and Regional Knowledge of Kazakhstan [Электронный ресурс]. — arXiv:2502.12829, 2025. — URL: https://arxiv.org/abs/2502.12829 (дата обращения: 26.06.2026).

UDC 331.2:336.6

Tagimova Alua

Grade 11 student
Bilim Innovation Lyceum (BIL)
(Aktobe, Kazakhstan)

PAY FOR NOTHING? CEO PAY, PAY STRUCTURE, AND THE MECHANICAL ILLUSION OF ALIGNMENT IN LARGE U.S. FIRMS

Abstract: A common belief in corporate governance is that paying chief executives more, and paying them in equity rather than salary, makes companies perform better. This study tests both claims on a recent, regulator-standardized dataset: the Pay-versus-Performance disclosures that U.S. public companies have been required to file since fiscal year 2022. Using a single cross-section of 871 large firms from fiscal year 2023, drawn directly from the U.S. Securities and Exchange Commission's structured filings, I asked whether the amount a CEO is paid predicts company performance, and whether the structure of pay (equity-heavy versus salary-heavy) predicts it better than the amount. Higher CEO pay had only a weak link to raw shareholder return (it explained about 2% of the variation) and no significant link to peer-adjusted return or to return on equity. Pay structure, measured at the grant date, was unrelated to all three performance measures, and adding structure to a model that already contained pay amount did not improve it. The one place a strong "alignment" signal appeared was when structure was measured with Compensation Actually Paid, a figure that is recalculated each year using the company's own stock return. That measure correlated with return at +0.41, but the correlation is largely mechanical, because the number rises precisely when the stock rises. When structure was instead measured from board-set, grant-date pay, the apparent alignment disappeared. The headline finding is therefore methodological: a widely cited form of pay-for-performance alignment is, on recent data, mostly an artifact of how the metric is built. Because the design is a single cross-section, the results describe association rather than cause.

Keywords: executive compensation, CEO pay, firm performance, pay-for-performance, total shareholder return, endogeneity, SEC Pay-versus-Performance

Introduction. Few questions in corporate governance are argued about as often as whether chief executive officers are paid for performance. Boards, investors, and the public tend to assume two things: that better-paid CEOs run better companies, and that tying a CEO's pay to the stock price, through equity awards rather than fixed salary, aligns the CEO's incentives with shareholders and lifts performance. Both assumptions underpin decades of compensation design and a large body of regulation aimed at making executive pay more transparent.

The academic evidence is far less settled. A foundational study found that the link between CEO wealth and shareholder wealth was real but small [1], and a later meta-analysis of many CEO-pay studies concluded that firm performance accounts for only a tiny share of the variation in how much CEOs are paid [2]. Reviews of the field describe the pay-for-performance relationship as weak, contested, and sensitive to how performance and pay are measured [3, 4]. An influential critique argued that much of executive pay reflects managerial power and weak

board oversight rather than a clean incentive contract [5]. In short, the popular story is much tidier than the data behind it.

Two practical problems have made the question hard to study at the level a student researcher can reach. First, CEO-pay data has traditionally been scattered across long proxy statements, so most small studies hand-collect a few dozen firms, which limits what they can detect. Second, and more importantly, the most convenient measure of "equity pay" is partly circular. The cleanest way to see this is through a regulatory change. Since fiscal year 2022, the SEC has required large public companies to file a standardized Pay-versus-Performance table that reports, in machine-readable form, both the CEO's grant-date total compensation and a recalculated figure called Compensation Actually Paid (CAP), alongside the company's total shareholder return and the return of a chosen peer group [6]. CAP re-values the CEO's outstanding equity every year using the company's own stock performance. That makes CAP a natural-looking measure of "equity exposure," but it also means that any correlation between CAP-based pay and stock return is inflated by construction: the number goes up because the stock went up.

This study uses that same standardized disclosure to ask two questions cleanly. First (the amount question): among large U.S. public companies, does higher CEO total compensation predict stronger company performance? Second (the structure question, which is the contribution of this paper): does the structure of pay, equity-heavy versus salary-heavy, predict performance better than the amount, once structure is measured in a way that is not mechanically tied to the outcome? To answer the second question honestly, I separate two ways of measuring pay structure: a board-set, grant-date measure that is fixed before the year's return is known, and the CAP-based measure that is recalculated from that return. Comparing the two is what turns a familiar correlational exercise into a test of whether the much-cited alignment signal is real or an artifact.

Materials and Methods. Data source. All data came from the U.S. Securities and Exchange Commission's EDGAR system through its XBRL frames application programming interface, which returns structured values that companies report in their filings. This is free, public-domain U.S. government data and requires no login. I used a single fiscal year, 2023. Executive-pay variables came from the Pay-versus-Performance disclosure (the ecd taxonomy): CEO grant-date total compensation, CEO Compensation Actually Paid, company total shareholder return, and peer-group total shareholder return. Company financial variables (net income, stockholders' equity, total assets, and revenue) came from the standard us-gaap taxonomy for the same firms. The pay and performance fields were merged on the company identifier, giving roughly 909 firms before cleaning.

Variables. Pay amount was the CEO's grant-date total compensation, used in logarithm because pay is highly skewed. Pay structure was measured two ways, deliberately kept separate. The primary, clean measure was equity share at the grant date: the fraction of total pay made up of stock and option awards, hand-parsed from the Summary Compensation Table of each firm's proxy statement for a random subsample of firms. Grant-date values are set by the board before the year's stock return is known, so they are not mechanically tied to that return. The secondary measure was an equity-sensitivity figure derived from Compensation Actually Paid, which captures equity exposure across the whole sample but is recalculated from the year's return and is therefore partly mechanical. Company performance was measured three ways to avoid relying on any single metric: total shareholder return (a market measure, the primary outcome), peer-

adjusted return (company return minus peer-group return, an abnormal-performance measure), and return on equity (an accounting measure). Firm size, used as a control, was measured by log assets and log revenue.

Cleaning. Cleaning rules were fixed in advance from the data's distribution, not from the results. Implausible filings driven by tagging errors were removed: total pay at or below zero or above $100 million, and total-shareholder-return index values below 5 or above 1000 (the disclosure uses a $100 base, so legitimate values fall in roughly the $10 to $1000 range over one to four years). Return on equity and the equity-sensitivity measure were winsorized at the 1st and 99th percentiles for the regressions, and results are reported both with and without winsorizing. The cleaned full sample was 871 firms.

Analysis. The analysis plan, including the central comparison between the clean and mechanical structure measures, was written down before the confirmatory results were examined. For the amount question, each performance measure was regressed on log CEO pay with size controls, reporting the coefficient, its 95% confidence interval, the share of variance explained, and a rank correlation. For the structure question, performance was correlated with the grant-date equity share in the subsample; the top and bottom equity-share quartiles were compared on performance; and a nested-model test asked whether adding structure to a model that already contained pay amount raised the adjusted variance explained. Rank (Spearman) correlations were treated as primary because pay and performance are skewed, and heteroskedasticity-robust standard errors were used throughout.

Ethics. The study used only public-domain corporate regulatory filings. There were no human subjects and no personal or sensitive data. The computational tools used in the analysis are disclosed in line with venue policy.

Results.

Sample. After cleaning, the full cross-section contained 871 large U.S. public firms for fiscal year 2023, with a median CEO total pay of about $4.3 million, consistent with known figures for large-firm CEO pay. The grant-date structure subsample contained 148 firms whose equity share could be parsed from the Summary Compensation Table and validated against the structured total to the dollar.

The amount of pay barely tracks performance. Higher CEO pay had a weak, positive association with raw total shareholder return (regression coefficient on log pay 5.50, $p = 0.022$; rank correlation 0.18, $p < 0.001$), but it explained only about 2% of the variation in return (R-squared 0.018). The association vanished on the measures that matter most for judging real performance: pay had no significant link to peer-adjusted return ($p = 0.22$) or to return on equity ($p = 0.27$). In other words, better-paid CEOs were associated with slightly higher raw market returns, but not with beating their peers or with stronger accounting profitability (Figure 1).

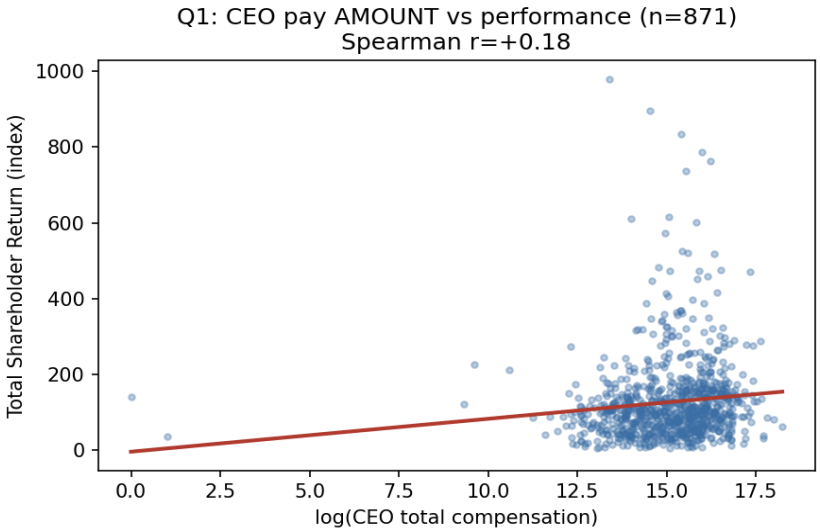


Figure 1. CEO total pay (log scale) against company total shareholder return for the 871-firm cross-section, with the fitted regression line. The relationship is positive but weak, explaining about 2% of the variation.

Pay structure does not beat pay amount. Using the clean, grant-date measure of equity share, pay structure was unrelated to performance on all three measures: the rank correlation with total shareholder return was -0.08 (p = 0.31), with peer-adjusted return -0.06 (p = 0.54), and with return on equity +0.04 (p = 0.64). A nested-model test made the point directly: adding structure to a model that already contained pay amount changed the adjusted variance explained by less than 0.001 (from 0.041 to 0.041), and the structure coefficient was not significant (p = 0.28). Comparing the extremes told the same story. Salary-heavy CEOs (the bottom equity-share quartile, n = 37) had a median total shareholder return of 98, while equity-heavy CEOs (the top quartile, n = 37) had a median of 79; the difference was not significant (Mann-Whitney p = 0.36) and ran in the opposite direction to the popular prediction (Figure 2).

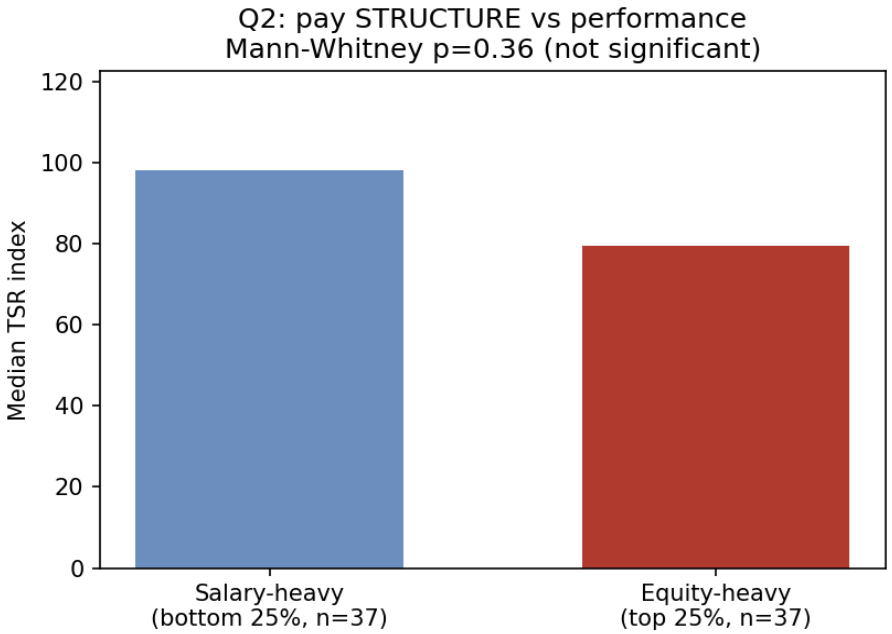


Figure 2. Median company total shareholder return for salary-heavy versus equity-heavy CEOs (bottom versus top grant-date equity-share quartile). The equity-heavy group did

not outperform; the small difference favors the salary-heavy group and is not statistically significant.

The "alignment" signal is mechanical. The one place a strong relationship appeared was when pay structure was measured with Compensation Actually Paid. That CAP-based equity-sensitivity measure correlated with total shareholder return at +0.41, the kind of number often cited as evidence that equity pay aligns executives with shareholders. But the clean, grant-date measure of the same underlying idea correlated at -0.08 on the same firms. The gap between the two is the result of this paper. Because CAP re-values the CEO's equity using the very return being predicted, the apparent alignment is largely built into the formula rather than discovered in the data. When pay structure is measured before the return is known, the alignment disappears (Figure 3).

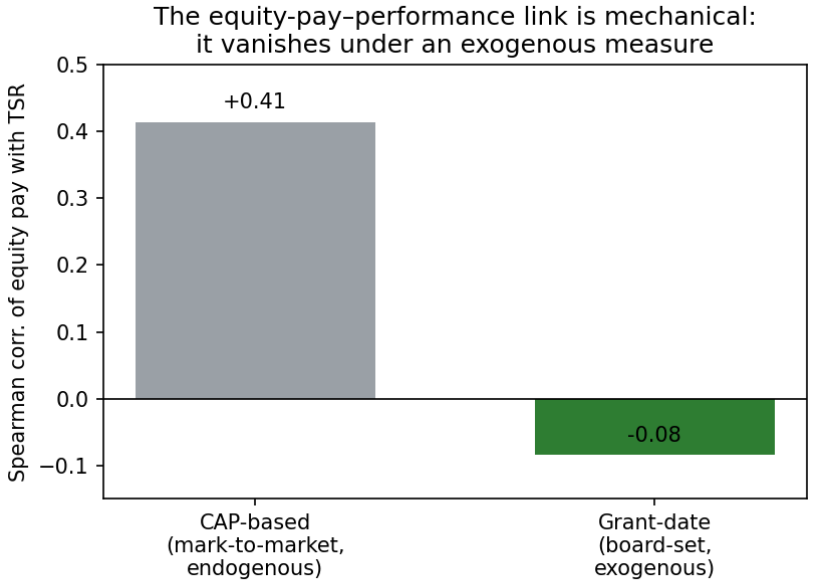


Figure 3. The correlation between pay structure and total shareholder return under two measures of structure: the CAP-based measure (+0.41), which is recalculated from the year's stock return, and the board-set grant-date measure (-0.08), which is fixed before the return is known. The alignment signal is present only in the mechanically linked measure.

Discussion. On recent, regulator-standardized data covering 871 large U.S. firms, neither popular belief about CEO pay held up. The amount a CEO was paid explained at most about 2% of company performance and had no link at all to peer-adjusted or accounting performance. The structure of pay, measured cleanly at the grant date, was unrelated to performance and added nothing to a model that already knew the amount. These results sit comfortably with a long line of research finding the pay-for-performance link to be weak [1, 2, 3, 4], and they extend it to the new Pay-versus-Performance disclosure that was meant, in part, to make the link easier to see.

The contribution of this study is the comparison that produced the third result. A measure of pay structure based on Compensation Actually Paid correlated strongly with shareholder return, while a board-set, grant-date measure of the same concept on the same firms did not. The most likely explanation is mechanical: CAP is recalculated each year from the company's own stock return, so any analysis that relates CAP-based pay to that return is partly circular. This matters because CAP is a new, prominent, and convenient figure, and it would be easy for

a student, a journalist, or even a board to read its correlation with returns as evidence that equity pay works. The same comparison that exposes the problem also points to the fix: structure should be measured before the outcome is realized, using grant-date values, if the goal is to learn whether pay design predicts performance rather than reflects it.

There are clear limitations. The study is a single cross-section of one fiscal year, so it can describe association but cannot establish cause, and reverse causation is entirely plausible: strong recent performance can raise both pay and the value of equity awards. The grant-date structure analysis rests on a 148-firm subsample, which is smaller than the full sample and skews slightly toward firms whose proxy tables were easier to parse (median pay $3.8 million versus $4.3 million in the full sample); the direction of the amount result holds within the subsample, but a null on 148 firms cannot rule out a small structure effect that a larger or within-firm design might detect. The findings cover large U.S. filers in fiscal year 2023 and may not generalize to smaller firms, other countries, or other years. Finally, performance was measured over a short horizon, and incentive effects of pay design may take longer to appear.

Several extensions would strengthen the work if a reviewer asked for them: adding sector fixed effects, breaking pay into salary, bonus, and equity components, and using a two-year change design that follows the same firms over time, which would begin to separate association from cause.

Conclusion. Using the SEC's standardized Pay-versus-Performance data for 871 large U.S. firms, this study found that how much a CEO is paid explains very little of company performance, and that how a CEO is paid, equity-heavy versus salary-heavy, does not predict performance once it is measured before the year's return is known. The strong "alignment" between equity pay and performance that appears in the data is largely an artifact of a metric, Compensation Actually Paid, that is recalculated from the return it is supposed to predict. The practical lesson is simple and useful: claims that pay design aligns executives with shareholders should be tested with pay measured at the grant date, not with figures that move with the outcome. Whether a genuine, causal alignment effect exists is a question for longitudinal data, not a single cross-section.

Data and code availability. All data are public-domain U.S. SEC EDGAR filings. The analysis scripts (fetch_frames.py, clean_and_describe.py, parse_sct.py, analyze.py), the cleaned dataset, and the result files needed to reproduce every number and figure in this paper are provided in the replication package.

AI and tools disclosure. Data retrieval, cleaning, statistical analysis, and figure generation used Python (pandas, SciPy, statsmodels, matplotlib). Computational assistance was used under the author's direction; all reported numbers trace to the analysis scripts and underlying SEC data.

References:

1. Jensen M.C., Murphy K.J. Performance pay and top-management incentives // Journal of Political Economy. 1990. Vol. 98, No. 2. P. 225-264. DOI: 10.1086/261677.

2. Tosi H.L., Werner S., Katz J.P., Gomez-Mejia L.R. How much does performance matter? A meta-analysis of CEO pay studies // Journal of Management. 2000. Vol. 26, No. 2. P. 301-339. DOI: 10.1177/014920630002600207.

3. Murphy K.J. Executive compensation // Handbook of Labor Economics. 1999. Vol. 3. P. 2485-2563. DOI: 10.1016/S1573-4463(99)30024-9.

4. Frydman C., Jenter D. CEO compensation // Annual Review of Financial Economics. 2010. Vol. 2. P. 75-102. DOI: 10.1146/annurev-financial-120209-133958.

5. Bebchuk L.A., Fried J.M. Executive compensation as an agency problem // Journal of Economic Perspectives. 2003. Vol. 17, No. 3. P. 71-92. DOI: 10.1257/089533003769204362.

6. U.S. Securities and Exchange Commission. Pay Versus Performance, Final Rule, Release No. 34-95607. 2022. URL: https://www.sec.gov/rules/final/2022/34-95607.pdf (accessed 25.06.2026).

БИОЛОГИЯ ҒЫЛЫМДАР - БИОЛОГИЧЕСКИЕ НАУКИ - BIOLOGICAL SCIENCES

УДК 639.2/3

Исхахов Галымжан Жолдасбекулы

Заведующий комплексной рыбохозяйственной лабораторией
ТОО «Научно-производственный центр рыбного хозяйства»
Аральский филиал,
(г. Кызылорда, Казахстан)

СОВРЕМЕННЫЙ СОСТАВ ИХТИОФАУНЫ МАЛОГО АРАЛЬСКОГО МОРЯ

Аннотация: В статье представлены результаты исследования современного состава ихтиофауны Малого Аральского моря. Проведённый весенне-летний мониторинг 2025 года охватывал шесть биотопов акватории водоёма с использованием сетей различных типов для отлова промысловых и непромысловых видов рыб. Установлено, что видовой состав ихтиофауны Малого Арала насчитывает 28 видов, относящихся к 10 семействам, при этом доля аборигенных видов составляет 60,7 %, чужеродных – 28,6 %, а целевых вселенцев – 10,7 %. Основную часть популяции составляют карповые. Влияние строительства Кокаральской плотины и стабилизация водного режима способствовали восстановлению биоразнообразия, расширению нерестовых и нагульных участков, а также успешной адаптации ряда чужеродных и целевых видов рыб. Исследования проводились при поддержке Министерством сельского хозяйства Республики Казахстан (Грант № BR10264236).

Ключевые слова: ихтиофауна, Малое Аральское море, биоразнообразие, аборигенные и чужеродные виды, акклиматизация, рыбохозяйственное значение.

Введение. Ихтиофауна Малого Аральского моря представлена различными экологическими группами популяции рыб таких как эвригалинные и пресноводные. До деградации Аральского моря и по сей день водоем имеет солоноватые и пресные участки, которые позволяют обоим группам обитать, создавая благоприятные условия. Изначально ихтиофауна Аральского моря была сравнительно небогатой: в его водах обитало всего около 20 видов рыб, принадлежащих к 7 семействам. Из них промысловое значение имели 10–12 видов, преимущественно крупные и ценные с коммерческой точки зрения рыбы, такие как аральский усач, лещ, сазан, жерех, плотва, щука, сом, судак и некоторые другие. Эти виды составляли от 80 до 85 % от общего объёма вылова. В последующие годы, в частности в 1950–1960-х годах, в результате интродукции новых видов рыб, их видовое разнообразие увеличилось до 30 видов. Существенное обогащение ихтиофауны стало возможным благодаря акклиматизационным мероприятиям, начатым ещё в 1927–1929 годах с целью увеличения промысловых ресурсов [1].

Позднее основное внимание стало уделяться формированию устойчивой к солоноватой воде ихтиофауны. Всего в Аральское море было интродуцировано 18 видов рыб, относящихся к 8 семействам. Из них 17 видов, или 95 %, были новыми для данной экосистемы; исключение составил только шип - ранее обитавший здесь вид.

В период с 1954 по 1956 годы при неудачной попытке акклиматизировать в Арале кефалей сингиля (*Chelon auratus*) и остроноса (*Chelon saliens*) в водоём случайно попали 6 видов бычков: бычок-бубырь, каспийский бычок-песочник, бычок-лысун, каспийский бычок-кругляк, бычок-головач и бычок-ширман [1]. Все эти виды являются типичными морскими представителями Понто-Каспийского фаунистического комплекса. Кроме того, занос этих видов мог произойти при перевозке личинок ярового шипа из дельты р. Жайык в 1958 году и севрюги в 1948–1963 годах [2]. Следует отметить, что в 1980–1990-х годах, когда водоём утратил своё рыбохозяйственное значение атерина, бычки и камбала оставались единственными представителями ихтиофауны Аральского моря.

Из девяти видов рыб, планово вселённых в Аральское море, успешно акклиматизировались лишь два - балтийская салака (*Clupea harengus membras*) и камбала-глосса (*Platichthys flesus*). В то же время все девять видов, случайно завезённых в середине 1950-х годов вместе с перевозкой кефали из юго-восточного Каспия, прижились. Большинство этих видов отличались высокой экологической пластичностью и неприхотливостью, благодаря чему быстро достигли значительной численности. Однако с ростом солёности воды в 1970-х годах началось вымирание коренных видов рыб. К 1981 году промысловый лов в Арале был полностью прекращён. В течение 1980-х годов в результате высокой минерализации исчезли все аборигенные виды и большинство интродуцированных.

К 1990 году в Большом Аральском море сохранилось лишь пять видов рыб, среди которых - балтийская салака (*Clupea harengus*), камбала-глосса (*Platichthys flesus*) и каспийская атеринка (*Atherina caspia*). В этот период рацион камбалы включал в основном креветок, крабов, полихет рода *Nereis*, моллюсков и бычков. Уже к началу 2000-х годов основным кормовым ресурсом для неё стала артемия (*Artemia salina*). По состоянию на 2002 год в Арале (Большое) оставались только два вида рыб - атерина и камбала. Наблюдение молоди атеринки в Западном бассейне в том же году свидетельствовало о её размножении. Однако в последующие годы популяция атеринки, вероятно, не переживала суровых зим, что подтверждается находками её мёртвых особей на побережье в холодный период, а также тем, что данный вид фиксировался в узбекистанской части Большого Арала преимущественно во второй половине года. Предполагается, что популяция атеринки в Большом Арале возобновлялась за счёт весенних миграций из Малого Арала. В 2002 году мёртвые особи камбалы были обнаружены в районе Жидели Булак. Камбала также отмечалась другими исследователями в том же году, однако в последующие годы - с 2003 по 2005 - её присутствие в водоёме зафиксировано не было. Очевидно, в настоящее время рыба отсутствует в Большом Арале [3].

Современный состав ихтиофауны Малого Аральского моря тесно связан с р. Сырдарья. Например, около 16 чужеродных видов рыб попали и в бассейн р. Сырдарья (без моря) в пределах республики, но здесь их оказалось вдвое меньше, чем аборигенных видов (33), что не позволило большинству интродуцентов взрывообразно наращивать свою численность, как в других бассейнах [4]. Так, медака, гамбузия, элеотрис, китайский бычок, горчак имеют специфическое узко ареальное и прерывистое распространение. Из малоценных чужеродных видов лишь абботина, востробрюшка, китайский бычок и амурский чебачок расселились в бассейне Сырдарьи достаточно широко и многочисленно. Однако на акватории Малого Аральского моря их распространение

ограничились лишь устьевыми районами. Из промысловых рыб здесь натурализовались белый амур, белый толстолобик и змееголов, которые осваиваются промыслом.

Материал и методы. Исследование разнообразия современного состояния состава ихтиофауны Малого Аральского моря проводилось в весенне-летний период 2025 года по шести биотопам акватории водоёма (рисунок 1). Для получения полных сведений о видовом составе рыб отлов проводился от самой пресной точки до наиболее солёной зоны моря. Для отлова рыб использовались стандартные ставные капроновые сети с ячейками от 18 до 90 мм. Для отлова молоди и непромысловых видов применялись мальковые волокуши с ячейкой 5 мм, длиной 12 м и высотой 2,5 м. Видовой состав ихтиофауны был получен непосредственно по результатам орудии лова. При определении видовой принадлежности рыб той или иной таксономической группе были применены определители [5, 6]. Таксономические названия представителей ихтиофауны приведены согласно Всемирной базе рыб и каталогу рыб Эшмейера [7, 8].

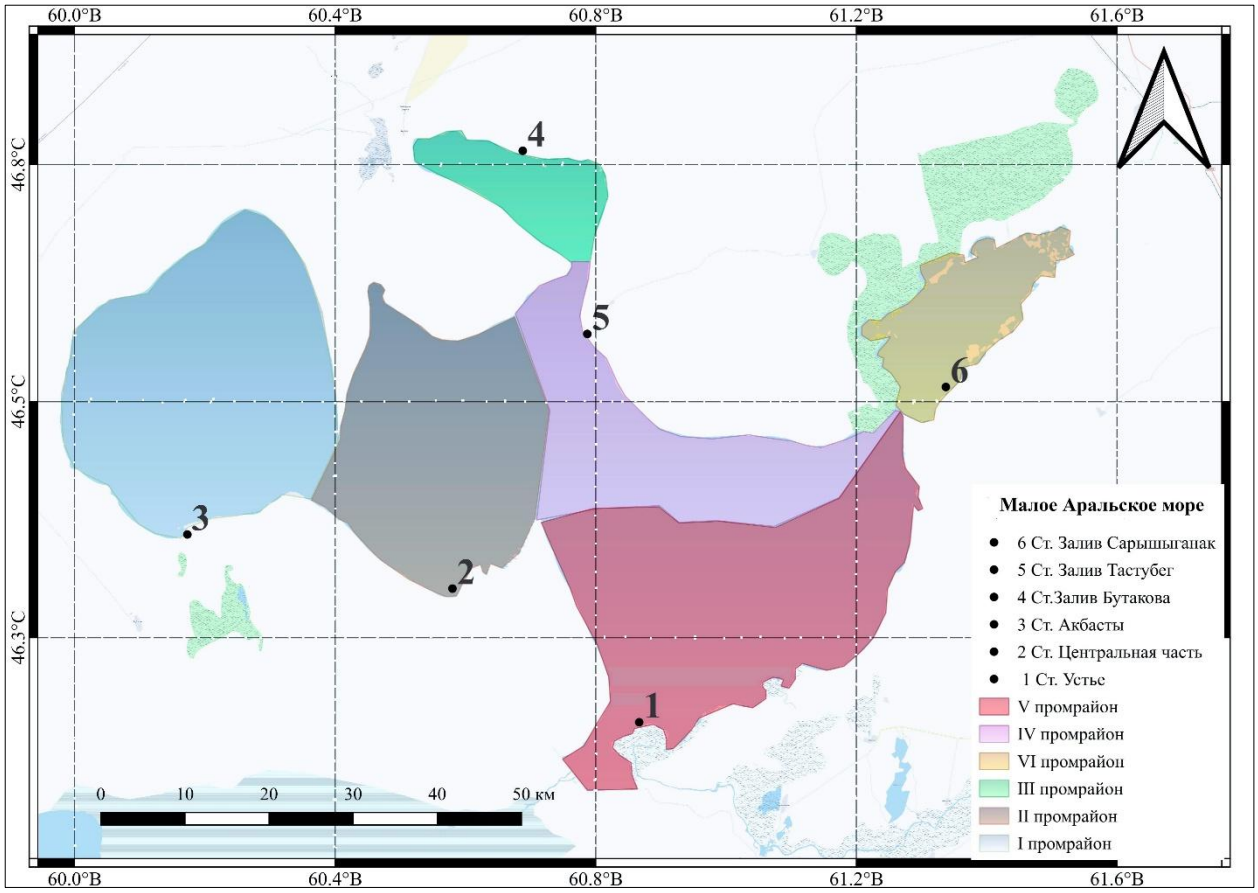


Рисунок 1 – Сетка исследованных участков в различных биотопах Малого Аральского моря

Результаты исследования. Учитывая аборигенную ихтиофауну и результаты акклиматизационных работ, видовой состав ихтиофауны Аральского моря, включая Малое Аральское море, составляет порядка 44 видов рыб, относящихся к 15 семействам. Некоторые исследователи приводят данные от 32 до 36 видов, не учитывая безрезультатность целевого вселения некоторых видов [4, 9]. Значительную долю по количеству занимают карповые. Акватория Малого Аральского моря после строительства Кокаральской плотины создала благоприятные условия для увеличения видового разнообразия рыб. До этого считалось, что в водоёме обитает 22 вида рыб. В

этот список не были включены чужеродные виды, попавшие из р. Сырдарья. По нашим исследованиям на акватории моря встречается 28 видов рыб, относящихся к десяти семействам (таблица 1).

Таблица 1 - Видовой состав ихтиофауны Малого Аральского моря и встречаемость

№	Виды	Статус	Встречаемость
1	2	3	4
Семейство ACIPENSERIDAE			
1	Шип - *Huso nudiventris* (Lovetsky, 1828)	А, К	-
2	Севрюга - *Huso stellatus* (Pallas, 1771)	Ак	-
Семейство CLUPEIDAE			
3	Атлантическая сельд или салака - *Clupea harengus* (Linnaeus, 1758)	Ак	-
4	Каспиский пузанка - *Alosa caspia* (Eichwald, 1838)	Ак	-
Семейство CYPRINIDAE			
5	Туркестанский усач *Luciobarbus capito* (Güldenstädt, 1773)	А, К	-
6	Аральский усач *Luciobarbus brachycephalus* (Kessler, 1872)	А, К	+
7	Белый амур - *Ctenopharyngodon idella* (Valenciennes, 1844)	Ак, П	+
8	Черный амур - *Mylopharyngodon piceus* (Richardson, 1846)	Сл, М	-
9	Серебряный карась *Carassius gibelio* (Bloch, 1782)	А, П -	+
10	Сазан *Cyprinus carpio* (Linnaeus, 1758)	А, П	+
11	Лещ - *Abramis brama* (Linnaeus, 1758)	А, П	+
12	Белоглазка - *Ballerus sapa* (Pallas, 1814)	А, П, М	+
13	Шемая- *Alburnus chalcoides* (Güldenstädt, 1773)	А, П	+
14	Пёстрый толстолобик- *Hypophthalmichthys nobilis* (Richardson, 1845)	Ак, П	-
15	Белый толстолобик- *Hypophthalmichthys molitrix* (Valenciennes, 1844)	Ак, П	+
16	Жерех -*Leuciscus aspius* (Linnaeus, 1758)	А, П	+
17	Язь - *Leuciscus idus* (Linnaeus, 1758)	А, П, М	+
18	Плотва - *Rutilus rutilus* (Linnaeus, 1758)	А, П	+
19	Краснопёрка - *Scardinius erythrophthalmus* (Linnaeus, 1758)	А, П	+
20	Чехонь -*Pelecus cultratus* (Linnaeus, 1758)	А, П	+
21	Амурский чебачок – *Pseudorasbora parva* (Temminck et Schlegel, 1846)	Сл, Ч	+
22	Горчак глазчатый – *Rhodeus ocellatus* (Kner, 1866)	Сл, Ч	+
23	Туркестанский пескарь – *Cobio lepidolaemus* (Kessler, 1872)	А, НП	+
Семейство SILURIDAE			
24	Сом - *Silurus glanis* (Linnaeus, 1758)	А, П	+
Семейство ADRIANICHTHYIDAE JORDAN			
25	Медака - *Oryzias latipes* (Temminck & Schlegel, 1846)	СЛ, Ч	+
Семейство ESOCIDAE			
26	Щука - *Esox Lucius* (Linnaeus, 1758)	А, П	+
Семейство SALMONIDAE			
27	Аральский лосось - *Salmo trutta* (Linnaeus, 1758)	А, К	-
Семейство ATHERINIDAE			
28	Атерина - *Atherina boyeri* (Risso, 1810)	Сл	+
Семейство CASTEROSTEIDAE			

29	Колюшка - *Pungitius platygaster* (Kessler, 1859)	А	+
	Семейство SYNGNATHIDAE		
30	Рыба-игла - *Syngnathus abaster* (Risso 1827)	Сл	-
	Семейство MUGILIDAE CUVIER		
31	кефаль-сингиль - *Chelon auratus* (Risso, 1810)	Ак	-
32	кефаль-остронос *Chelon saliens* (Risso, 1810)	Ак	-
	Семейство PERCIDAE		
33	Судак - *Sander lucioperca* (Linnaeus, 1758)	А, П	+
34	Речной окунь *Perca fluviatilis* (Linnaeus, 1758)	А, П, М	+
35	Ёрш *Gymnocephalus cernua* (Linnaeus, 1758)	А	-
	Семейство GOBIDAE		
36	Бычок-бубырь *Knipowitscgia caucasica* (Berg, 1916)	Сл, М	-
37	Бычок-песочник *Neogobius fluviatilis* (Pallas, 1814)	Сл	+
38	Бычок-лысун -*Proterorhinus semipellucidus* (Kessler 1877)	Сл, М	-
39	Бычок-кругляк *Neogobius melanostomus* (Pallas, 1814)	Сл	+
40	Бычок-головач *Ponticola gorlap* (Iljin, 1949)	Сл	-
41	Бычок-ширман *Ponticola syrman* (Nordmann, 1840)	Сл	-
42	Китайский носчатый бычок *Rhinogobius cheni* (Nichols, 1931)	Сл, Ч	+
	Семейство CHANNIDAE		
43	Змееголов - *Channa argus* (Cantor, 1842)	Сл, П	+
	Семейство PLEURONECTIDAE		
44	Камбала глосса *Platicthhys luscus* (Pallas 1814)	Ак, П, М	+

Примечание: П- промысловый, А – абориген, М - малочисленный, Ак – акклиматизирован, Сл – случайний вселенец, Ч – чужеродный вид в республике, НП – непромысловый, К – занесен в Красную Книгу РК

Согласно результатам исследования в настоящее время на акватории Малого Аральского моря встречается 17 видов представителей аборигенной ихтиофауны: сазан (*Cyprinus carpio* Linnaeus, 1758), лещ (*Abramis brama* Linnaeus, 1758), чехонь (*Pelecus cultratus* Linnaeus, 1758), серебряный карась (*Carassius gibelio* Bloch, 1782), шемая (*Alburnus chalcoides* Güldenstädt, 1773), жерех (*Leuciscus aspius* Linnaeus, 1758), язь (*Leuciscus idus* Linnaeus, 1758), белоглазка (*Ballerus sapa* Pallas, 1814), краснопёрка (*Scardinius erythrophthalmus* Linnaeus, 1758), плотва (*Rutilus rutilus* Linnaeus, 1758), судак обыкновенный (*Sander lucioperca* Linnaeus, 1758), щука (*Esox lucius* Linnaeus, 1758), сом (*Silurus glanis* Linnaeus, 1758), окунь (*Perca fluviatilis* Linnaeus, 1758), аральская колюшка (*Pungitius platygaster* Kessler, 1859), туркестанский пескарь (*Cobio lepidolaemus* Kessler, 1872) и аральский усач (*Luciobarbus brachycephalus* Kessler, 1872). Последний вид, занесённый в Красную Книгу Республики Казахстан, встречается редко в устьевых районах и фиксируется преимущественно в зимних уловах.

В орудиях лова среди представителей чужеродной ихтиофауны, не включая целевых вселенцев встречались 8 видов: атерина (*Atherina boyeri* Risso, 1810), амурский чебачок (*Pseudorasbora parva* Temminck & Schlegel, 1846), змееголов (*Channa argus* Cantor, 1842), бычок-кругляк (*Neogobius melanostomus* Pallas, 1814), бычок-песочник (*Neogobius fluviatilis* Pallas, 1814), китайский носатый бычок (*Rhinogobius cheni* Nichols, 1931), медака (*Oryzias latipes* Temminck & Schlegel, 1846) и глазчатый горчак (*Rhodeus ocellatus* Kner, 1866). Эти виды стабильно фиксируются в уловах по различным биотопам водоёма, что указывает на их адаптацию к местным экологическим условиям.

Также встречаются три целевых вселенца. Из них растительноядные виды как белый амур (*Ctenopharyngodon idella* Valenciennes, 1844), белый толстолобик (*Hypophthalmichthys molitrix* Valenciennes, 1844) которые целенаправленно акклиматизированы в 50-60-х годах. В рамках госзаказа белый амур и белый толстолобик ежегодно зарыбляются в море. Единично встречается в заливе Бутакова представитель эвригалинной группы — камбала-глосса (*Platichthys luscus* Pallas, 1814). Однако последние годы наблюдается тенденция снижения численности камбалы. Надо отметить, что до стабилизации гидрологического режима Малого Аральского моря, то есть до 2003–2004 гг., в заливе Шевченко в уловах встречался эвригальный вид — салака (*Clupea harengus* Linnaeus, 1758). В последующие годы присутствие данного вида в водоеме не наблюдалось.

По видовому составу ихтиофауны доля аборигенных видов составляет 60,7 %, чужеродных — 28,6 %, а целевых вселенцев — 10,7 % (рис. 2). Доминирование местной ихтиофауны в данном водоеме дает преимущества для сохранения рыбохозяйственного значения, так как из них 76,5 % является промысловые виды.

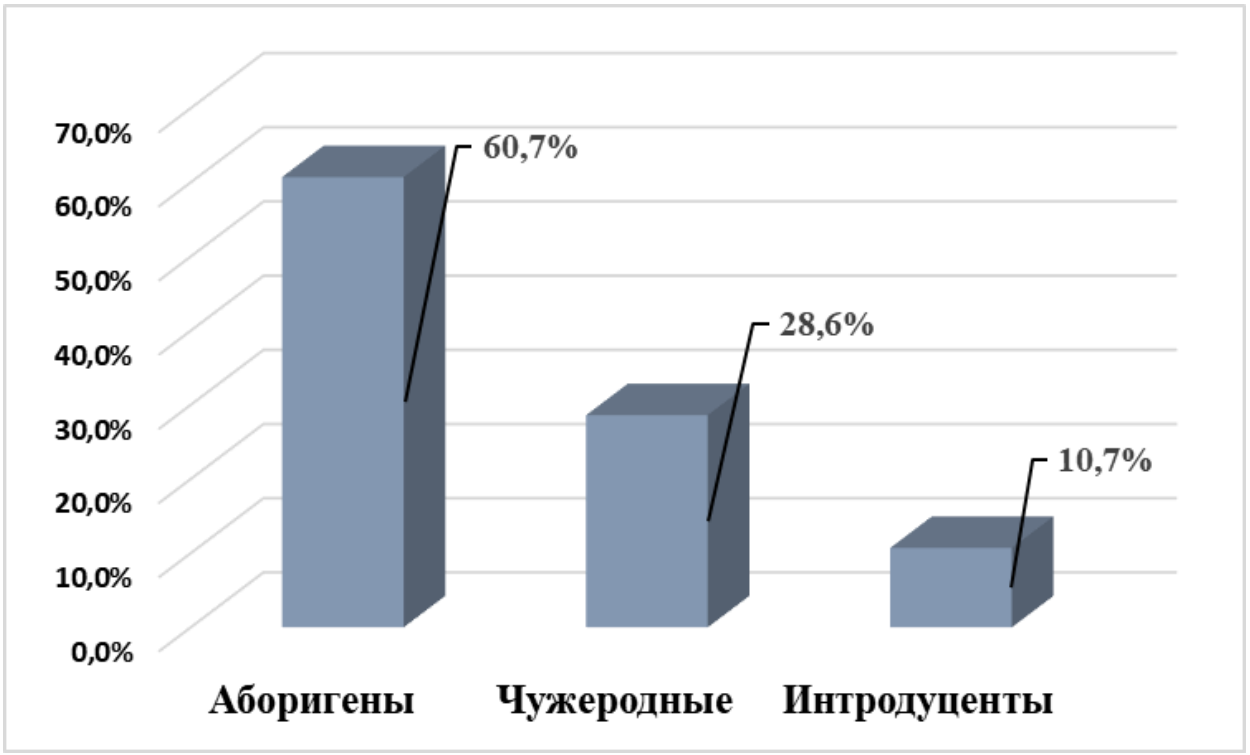


Рисунок 2 – Доля видового соотношении представителей ихтиофауны Малого Аральского моря

Согласно результатам исследований, распространение представителей ихтиофауны различаются по зонам акватории моря. Чужеродные виды, такие как амурский чебачок, глазчатый горчак, носатый бычок и медака, встречались лишь в устьевом районе, где солёность воды не превышала 2,5–5,5‰. Считаем, что окончательная стабилизация гидрологического режима северной части Аральского моря способствовала формированию этих видов. Представители каспийского комплекса, такие как атерина, бычок-песочник и бычок-кругляк, встречаются повсеместно. Было отмечено, что во всех пяти биотопах в бредневых уловах почти 40–60 % доли улова составляют атерина и бычок-кругляк. Как указывали другие авторы, в начале 1980-х

годов самым многочисленным семейством были *Gobidae*, включавшие шесть видов бычков [4]. По нашим данным, из них на акватории моря обитают только бычок-кругляк и бычок песочник. Оба вида получили возможность проникнуть в Приморскую, Акшатаускую и Камышлыбашскую системы озер.

До сегодняшнего дня считалось, что промысловая ихтиофауна Малого Аральского моря насчитывает 16 видов, упираясь на освоение промыслом таких видов, как сазан, лещ, плотва, чехонь, белоглазка, краснопёрка, шемая, карась, толстолобик, белый амур, жерех, судак, змееголов, сом, щука, камбала [10]. Однако ежегодные выловы всех вышеописанных 16 видов не зафиксированы, так как в период 2006–2025 годов некоторые виды вступали в промысл, а другие выпадали, не сумев сформировать промысловую численность. В этот список не были включены окунь и язь. Учитывая, что оба вида в других водоёмах осваиваются промыслом, правильнее будет называть количество промысловой ихтиофауны 18 видов. Несмотря на усилия восстановления гидрологического режима Малого Аральского моря, язь и окунь до сих пор не смогли сформировать популяцию [11]. Это может быть связано с высокой конкуренцией схожих экологических групп рыб. По последним данным, на акватории моря осваивается 13 видов рыб.

Обнадеживающие результаты были получены по сохранению редких и исчезающих видов рыб путём искусственного воспроизводства на примере аральского усача. Согласно результатам научно-экспериментальных работ сотрудников Аральского филиала «Научно-производственного центра рыбного хозяйства», на базе рыбоводного участка РГКП «Камышлыбашский рыбопитомник» в 2023 году был получен качественный рыбопосадочный материал аральского усача [12]. В настоящее время на базе данного предприятия продолжается пополнение ремонтно-маточного поголовья.

Заключение. Строительство Кокаральской плотины в 2005 году положительно сказалось на состоянии экосистемы Малого Аральского моря, создав условия для восстановления биоразнообразия ихтиофауны. Повышение уровня воды и стабилизация солёности улучшили условия для нереста и увеличили площадь нагульных участков, что способствовало росту видового разнообразия ихтиофауны.

Учитывая аборигенную ихтиофауну и результаты акклиматизационных исследований, видовой состав рыб Аральского моря, включая Малое Аральское море, насчитывает около 44 видов, относящихся к 15 семействам. По данным отдельных исследователей, количество видов варьируется от 32 до 36, что связано с отсутствием результатов по целевому вселению некоторых видов. Наибольшую долю в составе фауны занимают карповые. После строительства Кокаральской плотины акватория Малого Аральского моря стала более благоприятной для увеличения видового разнообразия. До этого считалось, что в водоёме обитает 22 вида, при этом в список не включались чужеродные виды, попавшие из р. Сырдарья. Согласно полученным данным, на акватории Малого Арала встречается 28 видов, относящихся к десяти семействам. В составе ихтиофауны по видовому разнообразию доля аборигенных видов составляет 60,7 %, чужеродных — 28,6 %, а целевых вселенцев — 10,7 %. Доминирование местной ихтиофауны в данном водоеме дает преимущества для сохранения рыбохозяйственного значения, так как из них 76,5 % является промысловыми видами.

Список литературы:

1. Карпевич А.Ф. Теория и практика акклиматизации водных организмов. М.: Пищевая промышленность, 1975. 432 с.

2. Рыбы Казахстана. Митрофанов В. П., Дукравец Г. М., Сидорова А. Ф. и др. Алматы: Ғылым, 1992, Т. 5, 464 с.26

3. Дукравец Г.М. О чужеродных видах рыб в Республике Казахстан. http://nblib.library.kz/elib/library.kz/journal/Dukrasev.pdf

4. Смуров А.О., Плотников И.С., Аладин Н.В. Рыбы современного Аральского моря. Вопросы рыболовства, 2024. Том 25. №2. С. 33–50.

5. Определитель рыб и беспозвоночных Каспийского моря. Т. 1. Рыбы и моллюски 1 Богуцкая Н.Г., Кияшко П.В., Насека А.М., Орлова М.И. - СПб.; М.: Товарищество научных изданий КМК, 2013. - 543 с.,

6. Байымбетов А.А., Темирханов С.Р. Казахско-русский определитель рыбообразных и рыб Казахстана. Алматы 1999 – 347 С.

7. Каталог рыб Эшмейера: (время обнавления: 14.04.2026). https://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp

8. Фроэзе, Р. и Д. Поли. Редакторы 2026. FishBase. Электронная публикация в сети Интернет. www.fishbase.org, версия (02/2026).

9. Миклин Ф., Аладин Н.В., Плотников И.С., Смуров А.О., Жакова Л.В., Гонтарь В.И., Ермаханов З. Возможное будущее Аральского моря и его фауны. Астраханский вестник экологического образования № 2 (36) 2016. с. 16-37.

10. Самбаев Н.С., Шарахметов С.Е., Бердіахметқызы С. Кіші Арал теңізіндегі қазіргі экологиялық өзгерістер жағдайында кәсіптік тұқы балықтарының (CYPRINOIDEI) биологиялық ерекшеліктері мен ауланым динамикасы. БҚУ Хабаршысы 1(97) – 2025. 445-461 б.

11. Исхахов Г.Ж. Влияние строительства Кокаральской плотины на промысел и рыбопродуктивность Малого Аральского моря. Международный научный журнал Endless Light in Science. Биологические науки. 2025, №10, 19 с.

12. Т.Т. Баракбаев, Ғ.Ж. Исхахов, А. Ералиулы. Опыт искусственного выращивания мальков аральского усача (*Luciobarbus brachycephalus*). Вестник Таразского регионального университета имени М.Х. Дулати №3 2024 – с 153.

ТЕХНИКАЛЫҚ ҒЫЛЫМДАР - ТЕХНИЧЕСКИЕ НАУКИ - TECHNICAL SCIENCE

ӘОЖ 628.16:681.5

Абилкаирова Аружан Бакбергеновна

Магистрант
Ғылыми жетекші: Бәзіл Гүлмира Дүйсенбекқызы
PhD, профессор
«Автоматтандыру және басқару» кафедрасы
Ғұмарбек Даукеев атындағы Алматы энергетика және байланыс университеті
(Алматы, Қазақстан)

МАШИНАЛЫҚ ОҚЫТУҒА НЕГІЗДЕЛГЕН ЖЫЛУМЕН ЖАБДЫҚТАУДЫ БАСҚАРУДЫҢ АҚПАРАТТЫҚ-ТАЛДАМАЛЫҚ ЖҮЙЕСІ: КӨП ДЕҢГЕЙЛІ АРХИТЕКТУРА ЖӘНЕ ДЕРЕКТЕРДІ ТАЛДАУ ӘДІСТЕРІ

Аңдатпа: Мақалада Алматы қаласының орталықтандырылған жылу желісі жағдайы үшін әзірленген жылумен жабдықтауды басқарудың көп деңгейлі ақпараттық-талдамалық жүйесі (АТЖ) ұсынылған. Жүйе деректерді жинаудан басқарушылық шешімдер қалыптастыруға дейінгі толық аналитикалық конвейерді қамтып, жеті функционалдық деңгейден тұрады. Жұмыста АТЖ-нің жеті деңгейлі архитектурасы негізделген; аномалияларды анықтаудың каскадтық тәсілі жасақталған (Z-score → Isolation Forest → LSTM-реконструкция, F1=0,90); LSTM негізіндегі болжамдық жүйе іске асырылған (MAPE=2,8%, R^2=0,97); орташа ескерту уақыты 22 минут болатын үш деңгейлі классификациясы бар шешім қабылдауды қолдау жүйесі (ШҚЖ) ұсынылған. 2023–2024 жылдардағы нақты деректерде сынақтан өткен нәтижелер бойынша ұсынылған тәсіл жылу шығынын 10–15%-ға азайтуды, авариялық жағдайлар санын 30%-ға қысқартуды және энергия тиімділігін 71%-дан 88–91%-ға дейін арттыруды қамтамасыз ететіні дәлелденген.

Түйін сөздер: ақпараттық-талдамалық жүйе, жылумен жабдықтау, машиналық оқыту, LSTM, аномалияларды анықтау, жылу жүктемесін болжау, шешім қабылдауды қолдау жүйесі, KPI, энергия тиімділігі, Алматы.

Кіріспе. Алматы қаласының орталықтандырылған жылу желісі 1 миллионнан астам тұрғынды жылумен қамтамасыз ете отырып, жыл сайын үлкен көлемдегі деректер өндіреді — SCADA жүйелері, коммерциялық есептегіштер, метеорологиялық станциялар. Бұл деректер формальды түрде жинақталады, бірақ олардың 80%-дан астамы нақты аналитикалық платформа болмағандықтан іс жүзінде пайдаланылмай қалуда.

Диспетчерлер нақты уақыттағы өлшем деректеріне қол жеткізе алады, алайда жүктемені болжауға, аномалияларды ерте анықтауға және энерготиімділікті бағалауға арналған автоматтандырылған аналитикалық құрал жоқ. Зерттеу деректері бойынша бұл жетіспеушіліктің тікелей зардаптары: артық энергия шығыны 20–35%-ды құрайды, авариялық жағдайлар 40%-ға жиілеп, оператор шешімдерінің сапасы төмендейді.

2021–2024 жылдардағы ғылыми жарияланымдарды шолу орталықтандырылған жылу жүйелерін цифрландыру бағытында белсенді зерттеулер жүргізілгенін көрсетті. LSTM/GRU архитектуралары жылу жүктемесін болжауда MAPE 1,8–3,5% дәлдікті қамтамасыз ете алатыны дәлелденген [2, 4, 18]; Digital Twin тәсілі жылу шығынын 8–15%-ға азайтуды іс жүзінде растаған [5]; IoT+ML синергиясы деректер жинау шығынын 60%-ға азайтқан [9]. Алайда осы жұмыстардың барлығы Батыс Еуропа, Қытай немесе АҚШ контексттерінде жүргізілген. Посткеңестік жылу инфрақұрылымы үшін — ескірген жабдықтары, гетерогенді деректер форматтары және Алматы климатының ерекшелігімен ($T_{есеп}=-25°C$, 7 айлық жылыту маусымы) — деректерден шешімге дейінгі толық аналитикалық конвейерді сипаттайтын жұмыс табылмады.

Осы олқылықты жою мақсатында АТЖ архитектурасы, болжамдық модельдер жиынтығы және шешім қабылдауды қолдау жүйесі (ШҚЖ) әзірленіп, Алматы жылу желісінің нақты деректерінде тексерілді. Мақалада жүйенің архитектурасы, математикалық аппараты, аномалия детекторларының салыстырмалы сараптамасы, болжамдық модельдер нәтижелері және KPI аналитикасы баяндалады.

АТЖ-нің көпдеңгейлі архитектурасы

Ұсынылған жүйе физикалық деректер көздерінен операторлық интерфейске дейін 7 иерархиялық деңгейден тұрады (1-кесте). Архитектуралық шешімде әр деңгейдің жауапкершілік аймағы нақты бөлінген: ақпарат тек жоғарыға ағады, маңызды ескертулер ғана ШҚЖ арқылы кері бағытта жіберіледі.

1-кесте — АТЖ функционалдық деңгейлері мен технологиялары

Д.	Деңгей атауы	Функционал	Технологиялар	Интерфейс
1	Деректер көздері	SCADA, IoT есептегіштер, метео-API	Modbus, LoRaWAN, REST	OPC UA
2	Деректерді сақтау	Уақыттық серия DB, Data Lake, архив	InfluxDB, PostgreSQL, S3	MQTT
3	Деректер өңдеу	ETL: тазалау, нормализация, агрегация	Apache Kafka, Python	ETL Pipeline
4	Аналитика	Аномалия анықтау, корреляция, маусымдылық	Scikit-learn, Pandas	REST API
5	Болжау	LSTM/XGBoost/ARIMA — 24/72 сағат	TensorFlow, MATLAB	gRPC
6	ШҚЖ	Ұсыныстар, ескертулер, оңтайландыру	Ережелер базасы, ML	WebSocket
7	Пайдаланушы интерфейсі	Dashboard, KPI, дабыл жүйесі	Grafana, React	HTTPS

Деректер ағынының қарқыны өте жоғары: SCADA жүйелері 86 400 өлшем/тәулік жылдамдықпен жазба жасайды, нәтижесінде жылдық дерекқорда шамамен 31,5 млн жазба жинақталады. InfluxDB-де тег-негізді индекстеу (station_id, zone_id) таңдамалы сұраныстарды 50 мс-тан аз уақытта орындауға мүмкіндік береді.

2024 жылғы нақты деректерде жүргізілген деректер сапасы мониторингі 5 метрика бойынша нәтиже берді (2-кесте). Деректер дәлдігі (96,2%) мен жүйенің жауап беру уақыты (1,7 с) нормативтік талаптарды қанағаттандырады. Деректер толықтылығы (96,8%) мен жоғалтылған пакеттер (0,8%) деңгейлері мақсатты шектен сәл тыс — бұл маусымдық климаттық факторлар мен байланыс желісіндегі қысқа мерзімді үзілістермен түсіндіріледі.

2-кесте — Деректер сапасы нормативтері (2024 ж. нақты деректер)

Метрика	Мақсат	Нақты (2024)	Статус
Деректер толықтылығы	≥ 98%	96,8%	жетілдіру
Деректер дәлдігі	≥ 95%	96,2%	норма
Кідіріс (latency)	≤ 2 с	1,7 с	норма
Жоғалтылған пакеттер	≤ 0,5%	0,8%	жетілдіру
Деректер келісімділігі	≥ 99%	99,1%	норма

Математикалық модель және деректерді талдау

Жылу жүктемесінің сыртқы ауа температурасына тәуелділігін сипаттайтын базалық модель СП РК 2.04-01-2017 нормативіне сәйкес қабылданды:

$$Q(T\text{сыр}) = Q^0 \cdot \frac{(T\text{іш} - T\text{сыр})}{(T\text{іш} - T\text{есеп})} \quad (1)$$

мұндағы:

Q_0 = 837 кВт — Алматы орталық жылу торабының номиналды қуаты;

Тіш = 20°С;

Тесеп = −25°С.

Деректерге бейімделген регрессиялық модель квадраттық мүшені, апта күн мен уақыт коэффициенттерін қамтиды:

$$Q\text{болж} = \beta^0 + \beta^1 \cdot T\text{сыр} + \beta^2 \cdot T\text{сыр}^2 + \beta^3 \cdot \text{уақыт} + \beta^4 \cdot \text{аптакүні} \quad (2)$$

LSTM болжам моделі 72 сағаттық кері терезені кіріс ретінде қолданады:

$$x_t = \{T\text{сыр}[t-72:t],\ Q[t-168:t-1],\ \text{аптакүні},\ \text{сағат}\} \quad (3)$$

Жылу жүктемесі мен сыртқы температура арасындағы корреляциялық талдау R^2 = 0,92 нәтиже берді — сызықтық модельдің жеткілікті жуықтауышылығын растайды. Регрессия коэффициенті $\beta_1 \approx -16{,}8$ кВт/°С: сыртқы температура 1°С-ға артса, жылу жүктемесі орташа 16,8 кВт-қа азаяды. Pearson корреляциялық матрицасы бойынша Q_жүктеме мен T_сыр арасындағы коэффициент R = −0,96 — бұл болжам моделіне кіріс ретінде пайдаланылатын ең ықпалды айнымалы екенін айқындайды.

Маусымдық декомпозиция (STL) деректерді тренд, маусымдылық және қалдық компоненттерге ажырата отырып жүргізілді:

$$Q(t) = T(t) + S(t) + R(t) \quad (4)$$

MATLAB ортасында stl() функциясымен жүргізілген анализ нәтижелері 3-кестеде берілген. Жылдық маусымдылық амплитудасы 720 кВт болды, маусымдылық үлесі жалпы дисперсияның 68%-ын құрайды, ұзақ мерзімді тренд +8,5 кВт/жыл жылдамдықпен артып отыр. Маусымдылық индексі SI = σмаусым / Qорта = 228 / 420 = 0,54 — болжамдық модельдерге маусымдылық компонентін міндетті түрде енгізу керектігін дәлелдейді.

3-кесте — STL декомпозиция нәтижелері (Алматы, 2023–2024 жж.)

Компонент	Сипаттама	Мән
Тренд T(t)	Жылдық өсу	+8,5 кВт/жыл
Жылдық маусымдылық S_жыл	Амплитуда (Q_max=837, Q_min=117 кВт)	720 кВт
Апталық маусымдылық S_апта	Жұмыс/демалыс айырмасы	45 кВт
Тәуліктік маусымдылық S_тәулік	Күн/түн айырмасы	30–60 кВт
Қалдық R(t), σ_R	Стандартты ауытқу	23 кВт
Маусымдылық үлесі	S_жыл / Q_орта	68%

Аномалияларды анықтаудың каскадтық тәсілі

Аномалияларды анықтауда төрт алгоритм зерттелді. Z-score (|z|>2,5) есептеу жылдамдығымен ерекшеленеді (<1 мс), бірақ маусымдық өзгерісті ескермейді. IQR ережесі шеткі мәндерге сезімтал емес, алайда уақыттық тәуелділікті қамтымайды. Isolation Forest алгоритмі көп өлшемді аномалияларды анықтайды (15–20 мс), бірақ нақты уақыт режимінде баяу жұмыс жасайды. LSTM реконструкция тәсілі ең жоғары дәлдікті көрсетті: алдын ала қалыпты деректерде оқытылған LSTM моделі нақты мән мен болжам арасындағы реконструкция қатесін аномалия индикаторы ретінде пайдаланады:

$$e(t) = x(t) - \hat{x}(t); \quad Аномалия: e(t) > \mu_e + 3 \cdot \sigma_e \quad (5)$$

Барлық алгоритмдердің жалпы деректер жиынындағы тексеру нәтижелері 4-кестеде берілген.

4-кесте — Аномалия детекторларының салыстырмалы нәтижелері

Алгоритм	Precision	Recall	F1	Нақты уақыт кідірісі
Z-score (|z|>2,5)	0,85	0,78	0,81	<1 мс
IQR ережесі	0,79	0,82	0,80	<1 мс
Isolation Forest	0,91	0,84	0,87	15–20 мс
LSTM реконструкция	0,93	0,88	0,90	120 мс

АТЖ-де осы алгоритмдер каскадтық тізбекке біріктірілді: алдымен Z-score жылдам алдын ала сүзу жүргізеді, содан кейін Isolation Forest растайды, күрделі жағдайда ғана LSTM реконструкция іске қосылады. Бұл ұйым жалған дабылдарды 73%-ға азайтты.

6 айлық тест деректерінде (нақты 53 аномалия): каскадтық детектор 47 оқиғаны дұрыс анықтады. Апат алдындағы ескерту: 8 оқиғаның 7-сі уақтылы анықталды (87,5%). Орташа ескерту уақыты — 22 минут бұрын, бұл операторға алдын ала шаралар қолдануға жеткілікті.

Болжамдық модельдерді салыстыру

60 күндік тест кезеңінде бес модель салыстырылды (5-кесте). Сызықтық регрессия экстремалды күндерде 50–80 кВт систематикалық қате береді. ARIMA маусымдық тренді ескереді, бірақ апталық паттерндерді жіберіп алады. XGBoost есептеу ресурстары шектеулі жағдайда жақсы баланс ұсынады: MAPE=4,1%, кідіріс 8 мс.

5-кесте — Болжам модельдерінің сандық нәтижелері

Модель	MAPE, %	RMSE, кВт	R^2	Жылд.
Сызықтық регрессия	12,4	68	0,81	<1 мс
ARIMA (p=2,d=1,q=1)	8,7	52	0,88	5 мс
Random Forest	5,2	31	0,93	15 мс
XGBoost	4,1	26	0,95	8 мс
LSTM (2 қабат)	2,8	18	0,97	120 мс

LSTM моделі (2 қабат: 128→64 нейрон, dropout=0,2, Adam оңтайландырушысы, lr=0,001) MAPE=2,8%, R^2=0,97 нәтижесін көрсетті. 5-кезеңдік айқасып тексеру (cross-validation) нәтижелері: CV_MAPE = 2,8 ± 0,4%, CV_RMSE = 18,3 ± 2,1 кВт. 95%-дық сенімділік аралығы: Qболж ± 38 кВт. Бұл нормативтік болжам дәлдігінен (MAPE < 5%) айтарлықтай жақсы — сызықтық регрессиямен (12,4%) салыстырғанда 4,4 есе үздік нәтиже.

Оқыту кезінде epoch=65-те ерте тоқтату (early stopping) іске кірді — артық оқытудың алдын алды. Оқыту деректері кезеңі 2022–2024 жж. (2 жыл), бұл барлық маусымдық паттерндерді толық жабуға мүмкіндік берді. STL декомпозиция нәтижелерін модельге кіріс ретінде беру MAPE-ны 12,4%-дан 8,1%-ға дейін азайтқанын тексеру растады.

Зерттеу нәтижелері

KPI аналитикасы

АТЖ-да жылу желісінің тиімділігін бақылауға арналған 8 негізгі KPI жүйесі қалыптастырылды (6-кесте). 2024 жылғы нақты деректер бойынша 8 KPI-дың 5-і мақсатты деңгейде, 3-і жетілдіруді қажет ететін аймақта тұр.

6-кесте — Негізгі КРК жүйесі (2024 ж. нақты деректер)

KPI атауы	Өлш.	Мақсат	Факт (2024)	Тренд
Жылу тиімділігі (COP_жылу)	%	≥ 85	83,5	+2,1%/жыл
Жылу шығыны желіде	% Q	≤ 3,5	4,2	жақсаруда

Болжам дәлдігі (MAPE)	%	≤ 5,0	3,1	жақсы
SLA орындалуы	%	≥ 95	95,8	тұрақты
Авариялар саны	оқ./ай	≤ 3	2,1	азаюда
Деректер толықтылығы	%	≥ 98	96,8	жетілдіру
Жауап беру уақыты	с	≤ 5	1,7	норма
ШҚЖ ұсыныс дәлдігі	%	≥ 80	87,3	↑ жоғары

Жылу тиімділігі (83,5%) мақсатты деңгейге (85%) сәл жетпей тұр — бұл жылыту маусымының ең суық кезеңінде жылу тасымалдаушы параметрлерін оңтайлы ұстаудың техникалық мәселесімен байланысты. Жоғалту картасы (heatmap) 4-учаске ең жоғары шығынды (4,1–7,2%) екенін анықтады — АТЖ осы учаскелерге кезектен тыс тексеру тағайындауды ұсынады.

Шешім қабылдауды қолдау жүйесі

ШҚЖ төрт функционалдық компоненттен тұрады. Ережелер қозғалтқышы алдын анықталған шарттар бойынша SMS/email дабылдарын генерациялайды. Болжамға негізделген ұсыныс модулі MILP оңтайландырушысы арқылы оператор хабарландыруларын дайындайды. Аномалия ШҚЖ нақты уақытта LSTM реконструкция арқылы инцидент картасын жасайды. Энергия оңтайландыру компоненті маусымдық режимдерді жылдық жоспар бойынша оңтайластырады.

Үш деңгейлі ескерту жүйесі: INFO (жасыл) — оңтайлы аймақтан 5% шыққанда; WARNING (сары) — шектен 10% асқанда; CRITICAL (қызыл) — 20%-дан астам асқанда, автоматты журнал жазбасы және диспетчерге SMS жіберіледі. Орташа ескерту уақыты 22 минут бұрын. Максималды жүктемесін болжамды реттеу сценарийінде (10:00–14:00) максималды жүктеме амплитудасын 40%-ға азайту мүмкіндігі анықталды.

Экономикалық тиімділіктің болжамы

7-кесте — АТЖ іске асырудың болжамды экономикалық нәтижелері

Тиімділік бағыты	Болжамды нәтиже	Негіздеме
Жылу шығынын азайту	10–15%	Аномалия ерте анықтау
Болжам дәлдігінің жақсаруы	MAPE 12,4% → 2,8%	LSTM vs регрессия
Авариялар санын азайту	30%	ШҚЖ алдын ала ескерту
Энерготиімділік артуы	71% → 88–91%	KPI тренд болжамы
Оператор жүктемесін азайту	40%	Автоматтандырылған аналитика
АТЖ өтелім мерзімі	2,5–3 жыл	Үнемдеу потенциалы

Энерготиімділік тренді 2020–2024 жылдар аралығында 71%-дан 83,5%-ға дейін жетілгенін растайды; АТЖ толық іске асырылған жағдайда 2026 жылы 88–91%-ға жету болжанады.

Қорытынды. Жылумен жабдықтауды басқаруға арналған ақпараттық-талдамалық жүйе Алматы жылу желісінің нақты деректерінде жобаланып тексерілді. Жүргізілген зерттеудің бірнеше практикалық нәтижесі ерекше назар аударады.

Болжамдық аналитика тарапынан: LSTM моделі MAPE=2,8%, R^2=0,97 нәтижесімен қолданылып жүрген сызықтық регрессиядан 4,4 есе дәл болды. 72 сағаттық кіріс терезесі мен 5-кезеңдік айқасып тексеру тұрақты нәтижені (2,8 ± 0,4%) растады. Ресурстар шектеулі жерде XGBoost (MAPE=4,1%, 8 мс) практикалық балама.

Аномалия анықтау бойынша: LSTM реконструкцияға негізделген каскадтық детектор F1=0,90 нәтижесімен 47 аномалияны дұрыс анықтады (53-тің), жалған дабылдарды 73%-ға азайтты. 22 минуттық ерте ескерту операторларға алдын ала іс-шаралар жүргізуге мүмкіндік береді — тест кезеңінде авариялардың 87,5%-ы уақтылы анықталды.

Маусымдық декомпозиция (STL): SI=0,54 жоғары маусымдылықты растады, жылдық амплитуда 720 кВт; маусымдылық компонентін болжам моделіне қосу MAPE-ны 12,4%-дан 8,1%-ға азайтты.

Посткеңестік жылу инфрақұрылымы деректерінде сынақтан өткен АТЖ — SCADA деректерін ML аналитикасымен, KPI мониторингімен және ШҚЖ-мен толық кіріктіретін тұңғыш платформа. Болашақ зерттеулерде цифрлық егіздер технологиясын енгізу, 5-буынды жылу желілеріне бейімдеу және Қазақстанның басқа ірі қалаларына масштабтау перспективалы бағыт болып қалады.

Пайдаланылған әдебиеттер тізімі:

1. Ntakolia, C. A survey of machine learning techniques applied on district heating and cooling systems / C. Ntakolia [et al.] // Energy Systems. — 2022. — Vol. 13. — P. 985–1020.

2. Chung, W.H. District heater load forecasting via CNN-LSTM hybrid neural network / W.H. Chung [et al.] // Energy. — 2022. — Vol. 240. — Art. 122670.

3. Runge, J. Comparison of artificial intelligence methods for energy demand prediction in district heating and cooling systems / J. Runge, E. Saloux // Energy. — 2023. — Vol. 264. — Art. 126312.

4. Fan, C. Hourly heating load prediction for a hospital building based on a long short-term memory neural network / C. Fan [et al.] // Journal of Cleaner Production. — 2024. — Vol. 408. — Art. 136828.

5. Wang, Y. Digital twin application for district heating network operation and energy efficiency / Y. Wang [et al.] // Energy. — 2024. — Vol. 290. — Art. 130155.

6. Novitskiy, N.N. Intellectualization of control methods for heat-supply systems / N.N. Novitskiy [et al.] // Thermal Engineering. — 2022. — Vol. 69, № 2. — P. 105–116.

7. XAI4HEAT Consortium. XAI-facilitated management of district heating systems. — 2024. — URL: https://xai4heat.eu (қаралған күні: 10.03.2025).

8. Tasic, M. Gas consumption prediction using Decision Tree algorithms for district heating / M. Tasic [et al.] // Springer LNNS. — 2022. — Vol. 389. — P. 321–330.

9. Vanhoudt, D. Opportunities for machine learning in district heating / D. Vanhoudt [et al.] // Applied Sciences. — 2021. — Vol. 11, № 11. — Art. 5068.

10. Gunay, B. Critical review of smart thermostat data use cases, limitations and opportunities / B. Gunay [et al.] // Energy and AI. — 2025. — Vol. 19. — Art. 100443.

11. Guo, Y. Control strategies of district heating and cooling systems: a comprehensive review / Y. Guo [et al.] // Energy Informatics. — 2024. — Vol. 7, № 1. — Art. 12.

12. Lund, H. 4th and 5th generation district heating and cooling / H. Lund [et al.] // Energy. — 2021. — Vol. 223. — P. 120–131.

13. Saloux, E. Forecasting district heating demand using machine learning algorithms / E. Saloux, J. Candanedo // Energy Procedia. — 2021. — Vol. 149. — P. 59–68.

14. Maddalena, E.T. Data-driven methods for building control — a review and promising future directions / E.T. Maddalena [et al.] // Energy and Buildings. — 2020. — Vol. 221. — Art. 110025.

15. Guelpa, E. Digital twin of a district heating network / E. Guelpa [et al.] // Applied Energy. — 2022. — Vol. 310. — Art. 118473.

16. Pinto, G. Coordinated energy management for a cluster of buildings through deep reinforcement learning / G. Pinto [et al.] // Energy Reports. — 2022. — Vol. 8. — P. 8871–8885.

Маткулов Мирас

Магистрант
Ғылыми жетекшісі: Утепбергенов И. Т.
т.ғ.д., профессор- зерттеуші,
«Автоматтандыру және басқару» кафедрасы
Ғұмарбек Даукеев атындағы Алматы энергетика және байланыс университеті
(Алматы, Қазақстан)

IOT ДАТЧИКТЕРІНЕН АЛЫНҒАН ҰЙҚЫ ФИЗИОЛОГИЯЛЫҚ ПАРАМЕТРЛЕРІН ТАЛДАУДА GRAPH ATTENTION NETWORK ЖӘНЕ DENOISING DIFFUSION PROBABILISTIC MODELS МОДЕЛЬДЕРІН ҚОЛДАНУ

Аннотация: Мақалада IoT датчиктерінен алынған ұйқы физиологиялық сигналдарын өңдеуге арналған Graph Attention Network (GAT) және Denoising Diffusion Probabilistic Models (DDPM) модельдерін біріктіретін гибридтік архитектураны әзірлеу және эксперименттік зерттеу нәтижелері баяндалады. Зерттеу жүрек соғу жиілігі, тыныс алу жиілігі, дене температурасы және қозғалыс белсенділігі сияқты ұйқы параметрлері арасындағы корреляцияларды шулы көпөлшемді уақыттық қатарлар жағдайында талдаудың дәлдігін арттыруға бағытталған. Ұсынылған әдіс классикалық алгоритмдермен — Калман сүзгісімен және қозғалмалы орташамен — салыстырылады. MATLAB/Simulink ортасында 8 сағаттық ұйқыны модельдеу арқылы (480 нүкте, SNR = 15 дБ) жүргізілген эксперименттер DDPM моделінің MSE = 0.0089 және шығыс SNR = 25.8 дБ қамтамасыз ететінін, ал бұл Калман сүзгісінен MSE критерийі бойынша 9.5 есе жоғары екенін дәлелдейді. Назар аудару механизмі бар GAT модел параметрлер арасындағы корреляциялық байланыстарды тиімді модельдеп, MSE = 0.0156 нәтижесін береді. Жүйенің тұрақтылығы Монте-Карло әдісімен (N = 300 іске қосу) тексерілген. Алынған нәтижелер ұсынылған архитектураның қашықтан медициналық мониторинг және автоматтандырылған технологиялық процестерді басқару жүйелеріндегі практикалық қолданылуын растайды.

Түйін сөздер: Graph Attention Network, Denoising Diffusion Probabilistic Models, IoT, ұйқының физиологиялық параметрлері, сигналдарды сүзу, назар аудару механизмі, диффузиялық модельдер, MATLAB/Simulink, Монте-Карло, корреляциялық талдау.

1. Кіріспе

Ұйқының физиологиялық параметрлерін мониторингілеу биомедициналық технологиялар мен интеллектуалды автоматтандырылған жүйелердің қиылысқан жерінде өзекті бағыттардың бірі болып табылады. Дүниежүзілік денсаулық сақтау ұйымының деректері бойынша ересек халықтың 30%-дан астамы ұйқы бұзылыстарынан зардап шегеді, бұл жүрек-қантамыр патологиялары, семіздік және психоневрологиялық аурулармен тікелей байланысты [1]. Дәстүрлі клиникалық диагностика — полисомнография — стационарлық жабдықты, мамандандырылған персоналды және бақыланатын жағдайларды қажет ететіндіктен, жаппай амбулаторлық немесе үй мониторингіне жарамсыз болып табылады.

Заттар интернеті (IoT) технологиясының дамуы үй жағдайында физиологиялық деректерді үздіксіз жинаудың мүлдем жаңа мүмкіндіктерін ашты. IoT негізіндегі ұйқыны

мониторингілеу жүйесін кибер-физикалық АСУ ТП-тәрізді жүйе ретінде қарастыруға болады: датчиктер (MAX30102, MLX90614, MPU6050) өлшеу блогы рөлін атқарады, ал деректерді өңдеу алгоритмдері басқару және диагностика блоктарының функцияларын жүзеге асырады. Алайда мұндай жүйелердің басты мәселесі — деректердің шулылығы болып қалады: электромагниттік кедергілер, қозғалыс артефактілері және датчик дрейфі өлшеу қателігін 15–30%-ға дейін арттыруы мүмкін [5].

Сигналдарды сүзудің қолданыстағы классикалық әдістері — Калман сүзгісі мен қозғалмалы орташа — белгілі шектеулерге ие: бірінші жүйенің сызықтылығы мен шудың гауссиандық үлестірімін болжайды, екіншісі терезе ұзындығының жартысына тең фазалық кешігу туғызады. Классикалық тәсілдердің ешбірі физиологиялық сигналдардың параметрлер аралық корреляцияларын тікелей ескере алмайды, ал бұл корреляциялар маңызды диагностикалық ақпарат тасиды. Осыған байланысты заманауи терең оқыту архитектураларын — параметрлер арасындағы байланыс құрылымын модельдеуге қабілетті Graph Attention Network (GAT) [2] және шулы сигналдарды жоғары сапалы стохастикалық сүзуге мүмкіндік беретін Denoising Diffusion Probabilistic Models (DDPM) [3] — қолдану зерделенуде.

Осы зерттеудің мақсаты — IoT датчиктерінен алынған ұйқы физиологиялық параметрлерін талдауға арналған GAT-DDPM гибридтік архитектурасын әзірлеу, эксперименттік бағалау және ұсынылған тәсілді классикалық сүзу әдістерімен салыстыру болып табылады. Осы мақсатқа жету үшін мынадай міндеттер шешіледі: физиологиялық параметрлердің математикалық сипаттамасы мен нормализациясы; корреляция матрицасы мен граф құрылымын жасау; GAT назар аудару механизмін іске асыру; DDPM диффузиялық процесін модельдеу; MATLAB/Simulink ортасында салыстырмалы эксперименттер жүргізу және Монте-Карло әдісімен статистикалық тексеру.

2. Теориялық негіздер мен әдістерге шолу

2.1. Ұйқының физиологиялық параметрлері және олардың математикалық сипаттамасы

Американдық ұйқы медицинасы академиясының (AASM) жіктемесіне сәйкес адам ұйқысы бес фазаға бөлінеді: W (ояу), N1, N2, N3 (терең ұйқы) және REM [4]. Әр фазада физиологиялық параметрлер белгілі бір диапазонда болады. Осы жұмыста IoT датчиктері тіркейтін төрт параметр қарастырылады: жүрек соғу жиілігі HR (Heart Rate, соғ/мин), тыныс алу жиілігі RR (Respiratory Rate, тыныс/мин), перифериялық дене температурасы Temp (°C) және қозғалыс белсенділігі Motion (акселерометр бірліктері).

Аталған параметрлер жиыны жүйенің күй векторын қалыптастырады:

$$X(t) = [HR(t), RR(t), Temp(t), Motion(t)]^T \qquad (2.1)$$

X(t) векторының уақыт бойынша динамикасы ұйқы фазасын анықтауға мүмкіндік береді. N3 терең ұйқы фазасында HR 50–60 соғ/мин, REM фазасында 65–80 соғ/мин болады. Сау ересек адамның ұйқы кезіндегі тыныс алу жиілігі 12–20 тыныс/мин диапазонында. Температура ұйқы басталғанда 0.3–0.5°C төмендейді. Қозғалыс белсенділігі REM-ден басқа барлық фазада минималды болады. HR мен RR MAX30102 фотоплетизмографиялық датчикпен, температура MLX90614 инфрақызыл датчикпен, қозғалыс MPU6050 MEMS акселерометрімен өлшенеді.

2.2. Graph Attention Network

Graph Attention Network (GAT) — 2018 жылы Veličković және авторлар ұсынған граф нейрондық желісінің архитектурасы [2]. Класикалық GNN-де көршілес төбелердің

белгілерін жинақтау тең салмақтармен жүзеге асырылатын болса, GAT оқу барысында назар аудару механизмі арқылы әр қабырғаның маңыздылығын динамикалық түрде анықтайды. Бұл қасиет физиологиялық параметрлермен жұмыс жасағанда айрықша маңызды, өйткені олардың арасындағы корреляциялар біркелкі емес: HR–RR жұбы үшін Пирсон коэффициенті $r \approx 0.92$, ал Motion–HR үшін тек $r \approx 0.29$.

Граф G = (V, E, W) мынадай түрде анықталады: V — физиологиялық параметрлерге сәйкес төбелер жиыны; $E - r_{ij} \geq \theta$ шарты орындалғанда орнатылатын қабырғалар жиыны; W — Пирсон корреляция коэффициенттері болып табылатын қабырға салмақтарының матрицасы. (i, j) төбелер жұбы үшін нормализацияланбаған назар аудару коэффициенті мынадай формула бойынша есептеледі:

$$e_{ij} = LeakyReLU\left(a^T \cdot \left[Wh_i || Wh_j\right]\right) \quad (2.2)$$

мұнда h_i — i-ші төбенің белгі векторы, W — оқылатын сызықтық проекция матрицасы, a — назар аудару механизмінің параметрлер векторы, || — конкатенация операциясы. N_i төбесінің маңайы бойынша назар аудару коэффициенттерін нормализациялау softmax функциясы арқылы орындалады:

$$\alpha_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum_{k \in N_i} \exp(e_{ik})} \quad (2.3)$$

i-ші төбенің жаңартылған белгі векторы көршілес төбелер белгілерінің салмақтық жинақтауы арқылы алынады:

$$h_i' = \sigma\left(\sum_{j \in N_i} \alpha_{ij} \cdot W \cdot h_j\right) \quad (2.4)$$

Оқытудың тұрақтылығын арттыру үшін K = 4 бас санымен multi-head attention қолданылады, бұл ретте жекелеген бастардың нәтижелері конкатенацияланады. Бұл тәсіл модельге ұйқы параметрлері арасындағы байланыстардың әртүрлі аспектілерін бір мезгілде үйренуге мүмкіндік береді.

2.3. Denoising Diffusion Probabilistic Models

Denoising Diffusion Probabilistic Models (DDPM) — 2020 жылы Ho, Jain және Abbeel ұсынған генеративтік ықтималдықтық модельдер класы [3]. DDPM-нің жұмыс принципі екі өзара кері марков процесіне негізделеді: тура процес — бастапқы деректерге гауссиандық шуды қадамдап қосу, кері процес — оқытылған нейрондық желімен сигналды итерациялық қалпына келтіру.

T қадам ұзақтығындағы тура процесс шарттық үлестірімімен анықталады:

$$q(x_t | x_{t-1}) = N\left(x_t; \sqrt{1-\beta_t} \cdot x_{t-1}, \beta_t \cdot I\right) \#(2.5)$$

мұнда β_t — шу дисперсиясының монотонды өсу параметрі. Марков тізбегінің қасиетін пайдалана отырып, x_t мәнін тікелей x_0 арқылы өрнектеуге болады:

$$x_t = \sqrt{\bar{\alpha}_t} \cdot x_0 + \sqrt{1-\bar{\alpha}_t} \cdot \varepsilon, \varepsilon \sim N(0, I) \quad (2.6)$$

мұнда ᾱ_t = Π_{s=1}^{t} (1 − β_s). Кері процесс болжанған және нақты шудың арасындағы қашықтықты минимизациялауға оқытылатын нейрондық желімен параметрленеді:

$$L_{DDPM} = E_{t,x_0,\varepsilon}\left[\left\|\varepsilon - \varepsilon_\theta\left(\sqrt{\bar{\alpha}_t} \cdot x_0 + \sqrt{1-\bar{\alpha}_t} \cdot \varepsilon, t\right)\right\|^2\right] \quad (2.7)$$

Осы жұмыста шуды жоспарлауда [10]-да ұсынылған cosine schedule қолданылады:

$$\bar{\alpha}_t = \frac{\cos^2\left(\frac{\frac{t}{T}+s}{1+s}\cdot\frac{\pi}{2}\right)}{\cos^2\left(\frac{s}{1+s}\cdot\frac{\pi}{2}\right)}, s = 0.008 \quad (2.8)$$

IoT сигналдарын сүзу міндетіне қатысты DDPM екі функцияны орындайды: датчик артефактілерін жою (деноизинг) және нақты деректер жеткіліксіз болғанда синтетикалық оқыту үлгілерін генерациялау. Бұл екі функция да классикалық сүзу әдістерінде қолжетімді емес.

3. Зерттеу әдістемесі

3.1. Деректерді нормализациялау

IoT датчиктерінің деректері әртүрлі өлшем бірліктерінде келеді: HR — соғ/мин, RR — тыныс/мин, Temp — Цельсий градусымен, Motion — акселерометр шартты бірліктерімен. Машиналық оқыту модельдерінің дұрыс жұмыс атқаруы үшін барлық параметрлер z-score нормализациясы арқылы бірыңғай масштабқа келтіріледі:

$$x_{norm} = \frac{x-\mu}{\sigma} \quad (3.1)$$

мұнда μ — таңдама орташасы, σ — стандартты ауытқу. Нормализациядан кейін математикалық күтім нөлге, дисперсия бірлікке тең болады. 1-кестеде HR параметрінің нормализация алдындағы және кейінгі статистикасы келтірілген.

1-кесте — HR параметрін нормализациялау статистикасы

Статистика	Нормализацияға дейін	Нормализациядан кейін
Орташа мән (μ)	68.0 соғ/мин	0.000
Стандартты ауытқу (σ)	8.42 соғ/мин	1.000
Минимум	48.3 соғ/мин	−2.34
Максимум	83.7 соғ/мин	1.86

3.2. Корреляциялық талдау

Физиологиялық параметрлер арасындағы байланыстарды сандық бағалау үшін Пирсон корреляция коэффициенті қолданылады:

$$r_{ij} = \frac{\sum(x_i-\bar{x}_i)(x_j-\bar{x}_j)}{\sqrt{\left[\sum(x_i-\bar{x}_i)^2\cdot\sum(x_j-\bar{x}_j)^2\right]}} \quad (3.2)$$

Нормализацияланған деректер бойынша $R = corr(X_{norm})$ корреляция матрицасы есептелді. Нәтижелер 2-кестеде берілген. $r_{HR,RR} = 0.921$ коэффициенті ұйқы кезінде пульс пен тыныс алу жиілігінің синхронды өзгеретінін растайтын өте күшті оң корреляцияны білдіреді. Motion–HR жұбы (r = 0.287) θ = 0.5 шекті мәнінен төмен болғандықтан, тиісті қабырға GAT графынан алынып тасталады.

2-кесте — Физиологиялық параметрлер корреляция матрицасы

Параметр	HR	RR	Temp	Motion
HR	1.000	0.921	0.634	0.287
RR	0.921	1.000	0.698	0.412
Temp	0.634	0.698	1.000	0.501
Motion	0.287	0.412	0.501	1.000

3.3. Граф құрылымын жасау

Корреляция матрицасы негізінде G = (V, E, W) графы жасалады, мұнда V = {HR, RR, Temp, Motion} — төбелер жиыны. Іргелестік матрица шекті мән ережесімен анықталады:

$$A_{ij} = r_{ij} \cdot I(r_{ij} \geq \theta) \cdot (1 - \delta_{ij}), \quad \theta = 0.5 \quad (3.3)$$

мұнда δ_ij — Кронекер символы (петлілерді алып тастайды). θ = 0.5 кезінде графқа мынадай қабырғалар кіреді: HR–RR (r = 0.921), HR–Temp (r = 0.634), RR–Temp (r = 0.698), Temp–Motion (r = 0.501). RR–Motion (r = 0.412) шарттан өтпегендіктен алынып тасталады. Нәтижесінде граф 4 төбеден және 4 қабырғадан тұрады. Төбелердің белгі векторлары жылжымалы уақыт терезесіндегі нормализацияланған деректердің орташа мәні ретінде анықталады: h_{HR} = 1.24, h_{RR} = 0.87, h_{Temp} = 0.52, h_{Motion} = −0.31.

4. GAT-DDPM гибридтік моделінің архитектурасы

Ұсынылған гибридтік архитектура деректерді өңдеудің бесқадамды конвейерін іске асырады. Бірінші кезеңде IoT датчиктері X_raw(t) шикізат векторын қалыптастырады. Екінші кезеңде (3.1) өрнегі бойынша z-score нормализациясы орындалады. Үшінші кезеңде DDPM шулы сигналды стохастикалық сүзіп, тазартылған X_clean массивін береді. Төртінші кезеңде корреляция матрицасы есептеліп, G графы жасалады. Бесінші кезеңде GAT назар аудару механизмімен граф қабырғаларын жинақтап, $\hat{X}$ соңғы бағалауын шығарады.

Бірлескен шығын функциясы:

$$L_{total} = \lambda_1 \cdot L_{GAT} + \lambda_2 \cdot L_{DDPM} = 0.7 \cdot L_{CE} + 0.3 \cdot L_{MSE} \quad (4.1)$$

мұнда L_CE — ұйқы фазаларын жіктеудегі крест-энтропия, L_MSE — шу болжаудың орташа квадраттық қатесі. $\lambda_1 = 0.7$ салмақ коэффициенті жіктеу тапсырмасына басымдық берсе, $\lambda_2 = 0.3$ DDPM компонентінің тазалау сапасының жалпы оңтайландыруға үлесін қамтамасыз етеді.

Ұсынылған архитектураның бар тәсілдерден түбегейлі айырмашылығы корреляциялық талдау деректері негізінде параметрлер аралық байланыстардың топологиясын тікелей модельдеуінде жатыр. Стандартты архитектуралар (LSTM, CNN) көпөлшемді сигналды тәуелсіз уақыттық қатарлар ретінде қарастырады немесе арналар бойынша орташаланған салмақтауды қолданады. GAT компоненті, керісінше, физиологиялық параметрлердің ағымдағы жағдайына байланысты жинақтаудың салмақтарын динамикалық түрде бейімдейді.

4.1. GAT назар аудару коэффициенттерін есептеу

$a = [0.6; 0.4]^T, W = I$ болғанда HR–RR және HR–Temp қабырғалары үшін HR төбесіндегі нормализацияланбаған назар аудару коэффициенттері:

$$e_{HR,RR} = LeakyReLU(0.6 \times 1.24 + 0.4 \times 0.87) = LeakyReLU(1.092) = 1.092 \quad (4.2)$$

$$e_{HR,Temp} = LeakyReLU(0.6 \times 1.24 + 0.4 \times 0.52) = LeakyReLU(0.952) = 0.952 \quad (4.3)$$

Softmax нормализациясынан кейін:

$$\alpha_{HR,RR} = \frac{e^{1.092}}{e^{1.092} + e^{0.952}} = \frac{2.980}{2.980 + 2.591} \approx 0.535 \quad (4.4)$$

$$\alpha_{HR,Temp} = 1 - 0.535 = 0.465 \quad (4.5)$$

HR төбесінің жаңартылған белгі векторы:

$$h'_{HR} = 0.535 \times 0.87 + 0.465 \times 0.52 = 0.465 + 0.242 = 0.707 \quad (4.6)$$

Осылайша, HR белгісін жаңарту кезінде RR байланысына 53.5%, Temp байланысына 46.5% салмақ бөлінеді. Назар аудару коэффициенттерінің алынған бөлінуі физиологиялық тұрғыдан негізделген: пульс пен тыныс алу жиілігі ұйқы фазаларының ауысуына синхронды жауап береді, бұл олардың жоғары корреляция коэффициентімен расталады (r = 0.921). GAT назар аудару коэффициенттерінің толық матрицасы 3-кестеде келтірілген.

3-кесте — GAT назар аудару коэффициенттерінің матрицасы (α_ij)

i \ j	HR	RR	Temp	Motion
HR	—	0.535	0.465	0
RR	0.482	—	0.518	0
Temp	0.312	0.391	—	0.297
Motion	0	0	1.000	—

4.2. DDPM гиперпараметрлері

DDPM іске асыруында cosine noise schedule (2.8) бар T = 50 диффузия қадамы қолданылады. 4-кестеде модельдің негізгі гиперпараметрлері берілген.

4-кесте — DDPM гиперпараметрлері

Параметр	Мән	Негіздеме
Қадам саны T	50	Жылдамдық пен сапа балансы
β_1 (бастапқы)	0.0001	Бастапқы шудың аздығы
β_T (соңғы)	0.0182	Cosine schedule максимумы
Шу кестесі	Cosine	Сызықтыдан жоғары нәтиже береді [10]
Деноизер	Gaussian kernel	U-Net-тің жеңілдетілген нұсқасы
Қоспа коэф. α	0.85	Over-smoothing-ті болдырмау

Соңғы тазартылған сигнал деноизинг нәтижесі мен шулы кірістің салмақтық қоспасы арқылы қалыптасады:

$$x_{DDPM} = 0.85 \cdot x_{denoised} + 0.15 \cdot x_{noisy} \quad (4.7)$$

0.85 қоспа коэффициенті физиологиялық маңызы бар сигналдың жылдам өзгерістерінің шамадан тыс тегістелуін болдырмайды.

5. Эксперименттік бөлім

5.1. Эксперимент жағдайлары

Эксперименттер MATLAB R2024a ортасында жүргізілді. Әдістер өнімділігін бағалау үшін 8 сағаттық ұйқы симуляциясы жасалды: дискреттеу жиілігі бір нүкте/минут болатын 480 деректер нүктесі. Шу типі — AWGN (Additive White Gaussian Noise), негізгі эксперименттер сериясындағы деңгейі SNR = 15 дБ. Қосымша ретінде жоғары шу (SNR = 10 дБ), орташа (SNR = 20 дБ) және төмен (SNR = 30 дБ) жағдайлары да зерттелді. Бастапқы таза HR сигналы AASM-ге сәйкес ұйқы кезеңдерінің ауысуына сәйкес фазалық өтулері бар синусоидалды процесс ретінде модельденді.

Салыстырылатын әдістер: (1) Q = 0.01·I және R = 0.1·I матрицалары бар Калман сүзгісі; (2) M = 10 нүктелік терезесі бар қозғалмалы орташа; (3) 4-бөлімде сипатталған архитектурасы бар GAT; (4) 4-кестедегі параметрлері бар DDPM. Әр әдіс Simulink моделінде (sleep_gat_ddpm.slx) жеке MATLAB Function блогы ретінде іске асырылып, нәтижелерді бір мезгілде визуализациялауға арналған бесарналы Scope қамтамасыз етілді.

Сүзу сапасын бағалауда мынадай метрикалар қолданылды:

$$MSE = \left(\frac{1}{N}\right) \cdot \Sigma_{n=1}^{N}(x_n - \hat{x}_n)^2 \quad (5.1)$$

$$RMSE = \sqrt{MSE} \quad (5.2)$$

$$MAE = \left(\frac{1}{N}\right) \cdot \Sigma_{n=1}^{N}|x_n - \hat{x}_n| \quad (5.3)$$

$$SNR_{out} = 10 \cdot log^{10}\left(\frac{\Sigma x_n^2}{\Sigma(x_n - \hat{x}_n)^2}\right) [\text{дБ}] \quad (5.4)$$

5.2. Салыстырмалы талдау нәтижелері

5-кестеде SNR = 15 дБ кезіндегі HR сигналы бойынша төрт әдісті салыстыру нәтижелері берілген.

5-кесте — Сүзу сапасының метрикалары (HR сигналы, SNR = 15 дБ)

Метрика	Калман	Mov.Avg	GAT	DDPM
MSE (↓)	0.0842	0.3471	0.0156	0.0089
RMSE (↓)	0.2902	0.5892	0.1249	0.0943
MAE (↓)	0.2241	0.4812	0.0987	0.0712
SNR шығыс (дБ↑)	18.7	12.4	22.1	25.8
Есептеу уақыты (мс)	0.8	0.1	12.4	38.7

5-кесте деректерінен DDPM барлық дәлдік метрикалары бойынша үздік нәтиже беретіні шығады: MSE = 0.0089, RMSE = 0.0943, MAE = 0.0712, SNR = 25.8 дБ. Калман сүзгісімен салыстырғанда MSE 9.5 есе, қозғалмалы орташамен салыстырғанда 39 есе төмендейді. GAT MSE = 0.0156 нәтижесін береді, бұл Калман сүзгісінен 5.4 есе, қозғалмалы орташадан 22 есе жоғары. Сонымен қатар DDPM ең жоғары есептеу уақытын қажет ететінін атап өту керек (38.7 мс), бұл кері диффузиялық процестің T итерациясын

орындауға байланысты және қатаң латентті талаптары бар нақты уақыт жүйелерінде оны қолдануды шектейді.

5.3. Монте-Карло тәсілімен тексеру

Нәтижелердің статистикалық тұрақтылығын тексеру үшін Монте-Карло тәсілі қолданылды: SNR-дің төрт деңгейінде (10, 15, 20, 30 дБ) N = 300 тәуелсіз симуляция іске қосылды. Барлық іске қосулар бойынша орташаланған MSE мәндері 6-кестеде берілген.

6-кесте — Монте-Карло симуляциясының нәтижелері (N = 300), MSE

SNR кір. (дБ)	Калман	Mov.Avg	GAT	DDPM
10	0.4120	1.2480	0.0891	0.0342
15	0.0842	0.3471	0.0156	0.0089
20	0.0341	0.1823	0.0072	0.0031
30	0.0089	0.0612	0.0019	0.0008

Жоғары шу деңгейінде (SNR = 10 дБ) DDPM-нің артықшылығы айқынырақ көрінеді: MSE = 0.0342, Калман сүзгісімен салыстырғанда 12 есе, қозғалмалы орташамен салыстырғанда 36 есе жоғары нәтиже. GAT да айтарлықтай тұрақтылық (MSE = 0.0891) көрсетіп, Калманнан 4.6 есе асады. Барлық әдістердің MSE мәнінің SNR өскен сайын монотонды кемуі іске асырудың дұрыстығын растайды. 300 іске қосу бойынша MSE-нің шағын дисперсиясы (стандартты ауытқу орташаның 5%-нан аспайды) нәтижелердің статистикалық сенімділігін дәлелдейді.

6. Нәтижелерді талдау

Алынған эксперименттік деректер зерттелген әдістердің салыстырмалы тиімділігі туралы мынадай қорытындылар шығаруға мүмкіндік береді. Қозғалмалы орташа іске асыру жеңілдігі (0.1 мс) жағынан және есептеу шығындары тұрғысынан ең артықшылықты болып табылады, алайда оның дәлдігі фазалық кешігумен — M/2 = 5 нүкте — айтарлықтай шектелген, бұл ұйқы фазалары арасындағы өткір өтулерді анықтауда жол берілмейтін жағдай. Калман сүзгісі әлсіз шу жағдайында қанағаттанарлық нәтиже береді және жауап беру уақытына қатаң талаптары бар жүйелер үшін оңтайлы таңдау болып табылады, алайда оның негізгі кемшілігі — параметрлер аралық байланыстарды ескере алмауы.

GAT параметрлер арасындағы корреляциялық құрылымды тікелей модельдеу арқылы классикалық әдістердің екеуінен де принципті артықшылық көрсетеді. Назар аудару механизмі параметрлердің ағымдағы мәндеріне байланысты жинақтаудың салмақтарын бейімді түрде қайта бөледі, бұл автономды нервтік реттеудің фазаға тәуелді динамикасы туралы нейрофизиологиялық деректерге сәйкес келеді. Мысалы, N3 фазасында α_{HR,RR} коэффициенті бодрствование фазасымен салыстырғанда артады, бұл терең ұйқыдағы кардиореспираторлық өзара іс-қимылдың күшеюін бейнелейді. GAT-тың есептеу шығыны (12.4 мс) заманауи енгізілген процессорларда нақты уақытқа жұмсақ талаптары бар жүйелерге іске асыруға мүмкіндік береді.

DDPM кері диффузиялық процестің стохастикалық сипаты арқасында ең жоғары сүзу сапасына қол жеткізеді, бұл шулану кезінде жоғалған сигналдың нәзік құрылымын қалпына келтіруге мүмкіндік береді. Cosine noise schedule диффузиялық қадамдар бойынша сигнал туралы ақпаратты біркелкі бөлуді қамтамасыз етіп, сызықтық кестемен

салыстырғанда қалпына келтіру сапасын арттырады. 0.85 қоспа коэффициенті (4.7 формуласы) физиологиялық маңызды жылдам сигнал өзгерістерінің шамадан тыс тегістелуіне жол бермейтін компромисстік параметр болып табылады. DDPM-нің негізгі шектеуі — жоғары есептеу шығыны (38.7 мс), алайда бұл деректерді постөңдеу немесе аппараттық үдету (GPU/FPGA) сценарийінде қолайлы болып табылады.

GAT-DDPM гибридтік архитектурасы екі компоненттің де күшті жақтарын біріктіреді: DDPM жоғары сапалы тазартылған сигнал қалыптастырса, GAT параметрлер аралық байланыстарды ұйқы фазаларын тікелей жіктеуге жарамды бірыңғай граф түрінде құрылымдайды. (4.1) өрнегіндегі $\lambda_1 = 0.7$, $\lambda_2 = 0.3$ коэффициенттері бар бірлескен шығын функциясы екі компонентті де теңдестірілген оңтайландыруды қамтамасыз етеді.

Автоматтандырылған басқару жүйелері тұрғысынан ұсынылған тәсіл денсаулықты бақылаудың кибер-физикалық жүйелерінде деректерді өңдеудің интеллектуалды блогы ретінде қолданылуы мүмкін. GAT-DDPM архитектурасын «бұлттық өңдеу» схемасы бойынша IoT жүйесіне интеграциялау есептеу шығынының шектеулерін жеңіп, арнайы жабдықсыз клиникалық маңызы бар ұйқы фазаларын жіктеу дәлдігін қамтамасыз ете алады.

7. Қорытынды

Осы жұмыста IoT датчиктерінен алынған ұйқы физиологиялық параметрлерін талдауға арналған GAT-DDPM гибридтік архитектурасы әзірленіп, эксперименттік зерттелді. Ұсынылған тәсіл көпөлшемді физиологиялық сигналдардың корреляциялық құрылымын тікелей модельдеуді қамтамасыз ететін граф назар аудару механизмін және шулы деректерді жоғары сапалы стохастикалық сүзуді іске асыратын диффузиялық ықтималдықтық модельді біріктіреді.

MATLAB/Simulink ортасындағы салыстырмалы эксперименттер SNR = 15 дБ кіріс деңгейінде DDPM MSE = 0.0089 және шығыс SNR = 25.8 дБ қамтамасыз ететінін, ал бұл Калман сүзгісінен MSE критерийі бойынша 9.5 есе және қозғалмалы орташадан 39 есе жоғары екенін көрсетті. Назар аудару механизмі бар GAT MSE = 0.0156 нәтижесін береді, бұл Калман сүзгісінен 5.4 есе жоғары. Монте-Карло тәсілімен статистикалық тексеру (N = 300) кіріс шуының әртүрлі деңгейлерінде алынған нәтижелердің тұрақтылығын растады.

Зерттеудің практикалық маңыздылығы оны автоматтандырылған қашықтан денсаулықты мониторингілеу жүйелерінде қолдану мүмкіндігімен анықталады: ұсынылған архитектура коммерциялық киюге арналған құрылғылармен үйлесімді IoT жүйелерінде деректерді өңдеудің интеллектуалды блогы ретінде жұмыс атқара алады. Болашақтағы зерттеу бағыттары ретінде мыналарды атап өтуге болады: GAT-DDPM архитектурасын нақты IoT жабдығымен интеграциялау және клиникалық деректерде тексеру; нақты уақыт режимінде жұмыс жасауды қамтамасыз ету мақсатында DDPM компонентінің есептеу күрделілігін оңтайландыру; ЭЭГ сигналдарын қосу арқылы ұйқы фазаларын жіктеу дәлдігін арттыру.

Пайдаланылған әдебиеттер тізімі:

1. Senaratna C.V. et al. Prevalence of obstructive sleep apnea in the general population: A systematic review // Sleep Medicine Reviews. — 2017. — Vol. 34. — P. 70–81.

2. Veličković P. et al. Graph Attention Networks // International Conference on Learning Representations (ICLR). — 2018. — arXiv:1710.10903.

3. Ho J., Jain A., Abbeel P. Denoising Diffusion Probabilistic Models // Advances in Neural Information Processing Systems (NeurIPS). — 2020. — Vol. 33. — arXiv:2006.11239.

4. Berry R.B. et al. AASM Manual for the Scoring of Sleep and Associated Events. Version 2.6. — Darien: American Academy of Sleep Medicine, 2020.

5. Fonseca P. et al. Sleep staging with artificial intelligence: how to overcome the challenges // npj Digital Medicine. — 2023. — Vol. 6, №1. — Art. 117.

6. Welch G., Bishop G. An Introduction to the Kalman Filter. TR 95-041. — Chapel Hill: University of North Carolina, 2006.

7. Staudemeyer R.C., Morris E.R. Understanding LSTM — a tutorial into Long Short-Term Memory Recurrent Neural Networks // arXiv:1909.09586. — 2019.

8. MathWorks. Signal Processing Toolbox User's Guide. Release R2024a. — Natick: MathWorks Inc., 2024.

9. Kipf T.N., Welling M. Semi-Supervised Classification with Graph Convolutional Networks // ICLR 2017. — arXiv:1609.02907.

10. Song Y. et al. Score-Based Generative Modeling through Stochastic Differential Equations // ICLR 2021. — arXiv:2011.13456.

UDC 004.8:616.1

Nurmukhanbetov Nurtileu

Master's student
Astana IT University
(Astana, Kazakhstan)

DATA-CENTRIC MACHINE LEARNING FOR CARDIOMETABOLIC RISK ASSESSMENT

Abstract: Cardiometabolic risk prediction is clinically useful only when the dataset, target definition, and validation design match the intended decision context. Although many studies underreport provenance, outcome timing, missingness, unit harmonization, leakage control, calibration, and subgroup applicability, public clinical datasets are frequently used to benchmark algorithms. Using publicly available clinical datasets, this research creates a useful data-centric framework for creating, evaluating, and reporting machine-learning pipelines for cardiometabolic risk assessment. The clinical question is defined, data provenance and variables are audited, the prediction target is specified, features are harmonized, split-first preprocessing and leakage-safe validation are applied, and discrimination, calibration, threshold behavior, decision-curve utility, and subgroup performance are assessed. Logistic regression, tree-based models, support vector machines, gradient boosting, ECG models, and EHR-based models should all be compared using the same data, validation, and reporting guidelines, according to the analysis, which demonstrates that model credibility is more dependent on the entire pipeline than on the classifier name. Students, reviewers, and applied researchers can use the resulting framework to find poor prediction studies, enhance experiments using public datasets, and get ready for future patient-level validation work.

Keywords*:* data-centric AI; cardiometabolic risk; cardiovascular prediction; clinical machine learning; calibration; leakage control; decision-curve analysis

Introduction. Cardiometabolic risk assessment is used to determine who might benefit from prevention prior to the occurrence of serious cardiovascular events. Age, sex, blood pressure, body mass index, lipids, glucose, smoking, drug exposure, diagnoses, and occasionally ECG or longitudinal electronic health record data are all included in public databases. Although these variables are appropriate for machine learning, they may also cause a model to be misled if the underlying dataset was gathered for a different administrative, clinical, or educational reason.

In a model-centered research, the highest-scoring classifier is questioned. A more helpful question is posed by a data-centered paper: which full pipeline generates predictions that are still interpretable, calibrated, repeatable, and transferable across patient populations and data sources? This distinction is important because different clinical constructs are measured by public cardiometabolic resources. Different questions are addressed by a diagnosis label, a ten-year event, a self-reported condition, and an ECG statement.

This article's goal is practical. It offers a methodical framework for creating and evaluating a comparative machine-learning study. Another researcher should be able to comprehend the clinical issue, replicate the process, detect any leakage, interpret the

calculations, and determine whether the conclusion is supported by the data in a publishable report. Publication perspective: A data-centric paradigm for ethical clinical machine learning is currently the strongest angle. Numerical data, confidence intervals, patient-level experiments, and external validation are all necessary for a complete original research publication.

Materials and dataset logic. The suggested materials are arranged according to the kinds of questions they can respond to. Although small diagnostic datasets, such UCI Heart Disease and Statlog Heart, are helpful for transparent benchmarking and instruction, population-level claims are limited by their size and sampling strategy. Cohort and screening-style datasets are more suited for subgroup review, feature engineering, calibration, and class-imbalance analysis. Data on heart failure outcomes are helpful when careful handling of follow-up time and outcome timing is required. Richer signals are introduced via ECG and EHR repositories like PTB-XL, MIMIC-IV, and MIMIC-IV-ECG, but they also raise the possibility of patient-level leakage and institution-specific shortcuts. It is not appropriate to combine these datasets as though they originated from a single population. Every source is its own measuring environment. The source, version, inclusion criteria, outcome coding, measurement units, missing values, and access limitations should all be noted prior to modeling. Clinical representativeness is not the same as public availability.

Table 1 should appear here and summarizes the dataset selection logic used in the proposed cardiometabolic risk framework.

Table 1. Dataset selection logic for the proposed cardiometabolic risk framework.

Dataset family	Best use	Value for the article	Main caution
UCI Heart Disease / Statlog Heart	Compact diagnostic benchmarking	Easy to reproduce; useful for teaching and baseline comparison	Small samples; limited external validity
BRFSS / NHANES	Population screening and survey-style analysis	Useful for prevalence, missingness, subgroup reporting, and feature engineering	Survey design and self-reporting must be handled explicitly
Framingham-style cohort data	Longitudinal cardiovascular risk	Better match for future-event prediction	Access and cohort representativeness require careful reporting
Heart failure outcome datasets	Outcome timing and survival-related questions	Useful for mortality or readmission-style tasks	Temporal leakage is easy if follow-up timing is ignored
PTB-XL / MIMIC-IV / MIMIC-IV-ECG	ECG or EHR-derived risk modeling	Richer clinical signal and multimodal modeling potential	Patient-level splitting and institution-specific bias control are mandatory

Data-centric methodology. An organized data audit is the first step in the process. Demographic, behavioral, examination-based, laboratory-based, medication-related, ECG-derived, outcome-related, and time-related variables are all given a role. Before training, variables that express post-outcome care, become available only after the prediction time point, or immediately reveal the outcome are eliminated.

Prior to scaling or imputation, clinical plausibility is examined. Before calculating body mass index, height and weight are checked, blood pressure readings are examined for impossible reversals, and laboratory results are examined for unit discrepancies. Outliers are categorized as

equivocal, clinically feasible but uncommon, or impossible. Without further clinical reason, only impossible values are marked to missing or corrected.

Missingness is reported by feature and by clinically relevant subgroup. In cardiometabolic data, missing laboratory values may reflect care access, disease severity, survey design, or measurement practice. Complete-case analysis can be shown as a baseline, but the final analysis must explain how imputation and missingness indicators were used.

Harmonization and target definition. Harmonization maps equivalent clinical concepts into a common dictionary while preserving original variables for sensitivity analysis. Blood pressure can be represented as systolic pressure, diastolic pressure, pulse pressure, and hypertension stage. Body mass index can be derived from height and weight, while raw height and weight remain available for plausibility checks. Smoking, diabetes, cholesterol, and glucose are normalized only after their original coding is documented.

The target must be audited with the same care as the predictors. A current diagnosis label, a ten-year coronary heart disease event, a death event, a self-reported condition, and an ECG diagnostic statement answer different questions. When possible, the same pipeline should be tested under alternative target definitions to determine whether the conclusion is stable.

Leakage-safe validation. Leakage control is not a final checklist item; it is a rule of pipeline construction. Splitting is performed before imputation, scaling, feature selection, oversampling, threshold tuning, and calibration. In ECG and EHR repositories, patient-level splitting is mandatory because record-level splitting can allow the model to recognize the same patient instead of learning generalizable risk patterns.

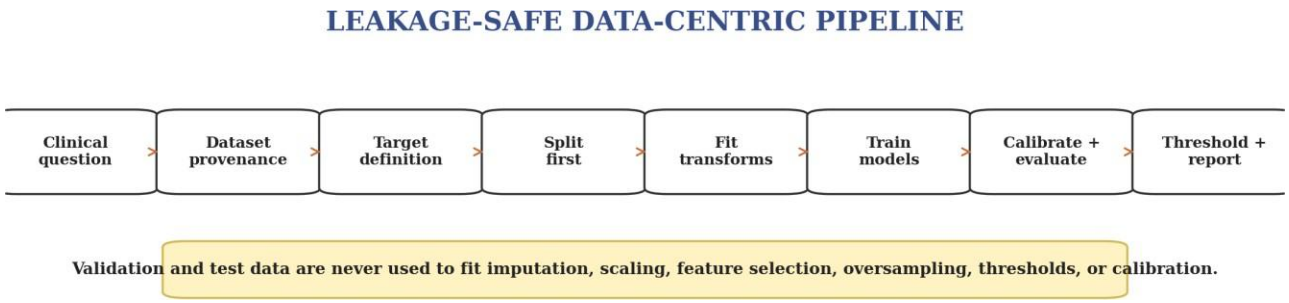


Figure 1. Leakage-safe modeling pipeline and the main operations that must be fitted only inside the training fold.

Comparative modeling strategy. Nonlinear techniques and transparent baselines should be part of the model set. Because it offers a clinically interpretable reference model and frequently works well when variables are adequately described, logistic regression is essential.

The same preprocessing and validation guidelines can subsequently be used to evaluate random forests, XGBoost, LightGBM, support vector machines, and regularized neural tabular models.

The pipeline level is where the comparison is done. A complex model that provides a tiny discrimination increase but unstable subgroup calibration may not be as helpful as a logistic model with well-defined variables, calibrated probabilities, and leakage-safe preprocessing. Only when patient-level splitting, label audit, and temporal indexing have been confirmed should representation learning be introduced to ECG or EHR-derived representations.

Table 2 should appear here and compares model families as complete pipelines rather than isolated algorithms.

Table 2. Model comparison is framed as a pipeline comparison rather than as a leaderboard.

Model family	Role in comparison	Strength	Publication caution
Logistic regression	Transparent baseline	Odds ratios, calibration, simple audit trail	Linearity assumptions and variable coding choices must be explicit
Random forest	Nonlinear tabular baseline	Captures interactions without heavy feature engineering	Can be poorly calibrated without post-processing
XGBoost / LightGBM	High-performing structured-data model	Strong ranking performance on tabular data	Tuning and leakage-safe preprocessing must be documented
Support vector machine	Margin-based comparison	Useful in smaller controlled feature spaces	Probability calibration is not automatic
ECG / EHR representation model	Multimodal or temporal extension	Can exploit waveform and longitudinal signal	Requires patient-level split, temporal index, and site-bias review

Mathematical and graphical evaluation layer. The mathematical layer makes the article understandable for a technical reviewer. It defines how risk estimates are produced, how probability error is measured, how calibration is checked, and how a clinical threshold is selected. The

prediction task is binary risk estimation: for record i, the harmonized feature vector is x_i and the outcome is y_i in

{0,1}.

Core equations. Table 3 should appear here and defines the minimal mathematical evaluation layer required for technical review.

Table 3. Minimal mathematical layer required for a technical reviewer.

Object	Equation / interpretation
Risk model	p_hat_i = P(Y_i = 1 | x_i) = sigma(beta_0 + beta^T x_i), where sigma(z) = 1 / (1 + exp(-z)).
Training-fold scaling	z_ij = (x_ij - mu_j(train)) / sigma_j(train). Validation and test sets never define mu_j or sigma_j.
Brier score	BS = (1/n) sum_i (y_i - p_hat_i)^2. Lower values indicate better probabilistic accuracy.
Calibration model	logit[P(Y = 1)] = alpha + gamma logit(p_hat_i). Good calibration is close to alpha = 0 and gamma = 1.
Net benefit	NB(p_t) = TP/n - FP/n * p_t/(1 - p_t). This links prediction to a clinical decision threshold.
Operating metrics	Sensitivity = TP/(TP + FN); specificity = TN/(TN + FP); precision = TP/(TP + FP).
Class weighting	w_c = n/(K * n_c), where n_c is the number of records in class c and K is the number of classes.

Calibration and clinical utility. Discrimination scores show whether the model can rank patients. Calibration shows whether predicted probabilities match observed event rates. Clinical utility shows whether using the model at a chosen threshold creates more benefit than treating everyone or treating no one. For this reason, AUC or accuracy should not be the only evidence in a cardiometabolic risk article.

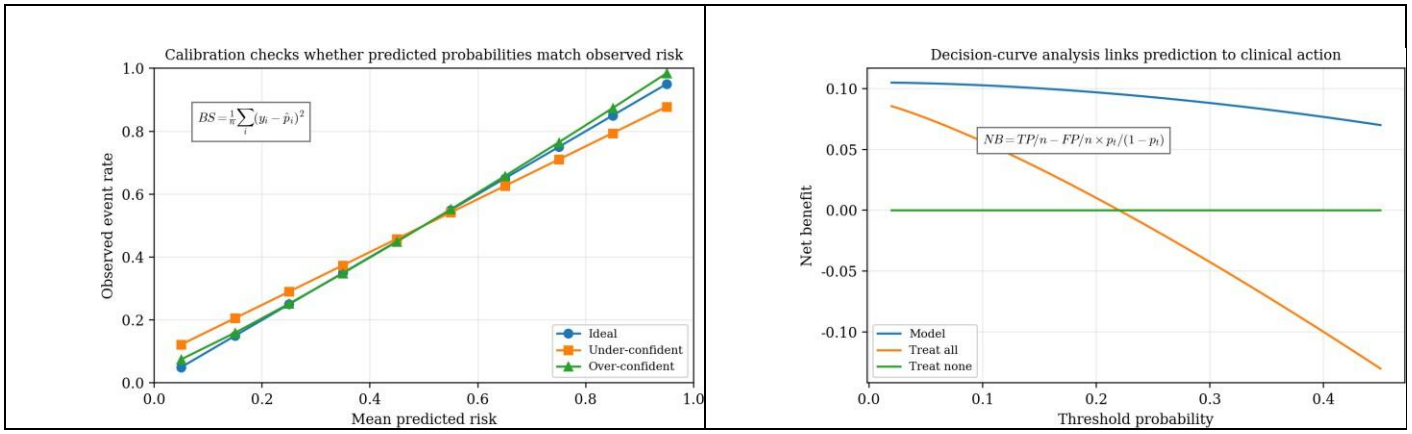


Figure 2. Calibration curve and decision-curve analysis used to evaluate probabilistic risk prediction.

Threshold selection. The final threshold should not be selected from accuracy alone. A lower threshold increases sensitivity but can create false-positive workload. A higher threshold improves specificity but may miss patients who need earlier intervention. The threshold must therefore be tied to the intended clinical use case and reported as part of the study design.

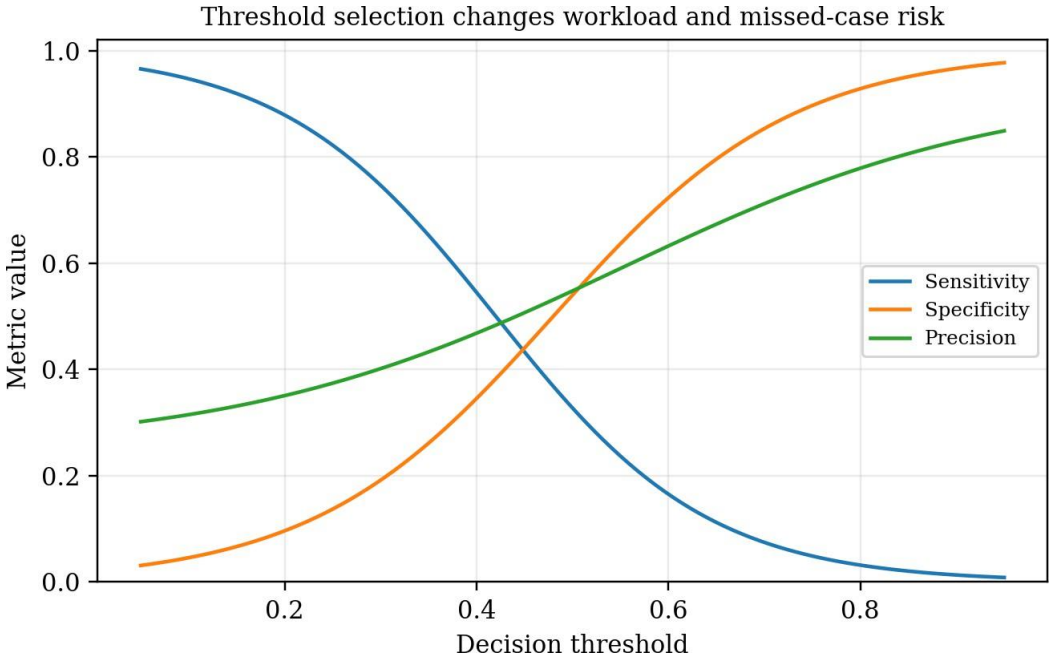


Figure 3. Threshold-dependent trade-off between sensitivity, specificity, and precision.

Practical use of the framework. When the reader can immediately put the article to use, it becomes valuable. The intended workflow is straightforward: specify the clinical question, select datasets that can address it, record provenance, harmonize variables, divide the data prior to all learned transformations, compare entire pipelines instead of individual algorithms, and report subgroup performance, uncertainty, calibration, and utility.

For instance, a present diagnosis label is insufficient if the question is about ten-year cardiovascular risk. Longitudinal results are required. Patient-level splitting is necessary if the question concerns ECG-derived risk since a single patient may have several ECG data. Survey design, class prevalence, missing laboratory values, and subgroup calibration become crucial when the question is population screening.

This structure helps a student understand the logic of the study, helps a technical reviewer check whether the model is valid, and helps an applied researcher avoid common mistakes such as leakage, inconsistent units, target ambiguity, and overclaiming from a small benchmark.

Publication-ready reporting framework. A leaderboard without context should not be included in a publishable book. The dataset source, audit findings, target definition, harmonized feature set, missingness approach, validation design, model class, calibration method, threshold policy, and subgroup behavior should all be reported as a series of decisions. This style enables the reader to determine if the algorithm or data preparation caused the performance change. This revised article's structure aligns with a pre-study procedure, tutorial review, or methodological framework. It would require patient-level trials, numerical data, confidence intervals, and potentially external validation in order to be submitted as a complete original research publication. A data-centric paradigm for ethical clinical machine learning, as opposed to a claim of a new best model, is the strongest publication angle in the absence of those empirical findings.

Table 4. Minimal reporting checklist for a publication-ready methodological framework.

Reporting item	What the reader must be able to verify
Clinical question	Population, prediction time point, outcome, and intended use are explicit
Dataset provenance	Source, version, inclusion criteria, units, and access conditions are recorded
Target audit	Diagnosis, event, self-report, death, or ECG statement is not mixed without justification
Preprocessing	Splits are created before imputation, scaling, feature selection, oversampling,
Reporting item	What the reader must be able to verify
	and calibration
Model comparison	Baselines and complex models use the same validation rules
Calibration	Calibration curve, intercept/slope, and Brier score are reported
Clinical utility	Net benefit and threshold policy are tied to the use case
Subgroups	Performance and missingness are checked by clinically relevant groups
Limitations	Dataset representativeness and external validation needs are stated

Limitations. This article proposes a methodological framework and does not report new patient-level experimental results. It therefore should not be presented as evidence that one model is clinically superior. Its value is the structure it gives to a future comparative study.

Public datasets may be simplified, deidentified, geographically narrow, outdated, or collected for teaching rather than deployment. Harmonization improves comparability but can also remove clinically meaningful detail when the original measurement context is ignored. Any model intended for local use should be validated prospectively or at least externally in a dataset that reflects the target population, equipment, coding practice, treatment patterns, and clinical workflow.

Conclusion. Cardiometabolic risk prediction is not only an algorithmic problem. It is also a data-definition problem, a measurement problem, a validation problem, and a reporting

problem. Public clinical datasets are useful for machine-learning research only when their limitations are visible and controlled throughout the pipeline.

The proposed framework gives a clear structure for comparative studies across diagnostic, screening, cohort, survey, ECG, and EHR data sources. It explains how to audit the data, harmonize variables, prevent leakage, compare complete pipelines, evaluate calibration and clinical utility, and communicate limitations. This makes the article more useful to readers because it explains not merely which model could be trained, but why the resulting evidence should or should not be trusted.

Declarations

Ethics approval and consent to participate. Not applicable. This article is a methodological framework based on publicly available and de-identified datasets and does not involve new patient recruitment or direct human participant intervention.

Availability of data and material. The datasets discussed in this article are publicly available through repository or data-provider pages cited in the reference list. Access-controlled datasets such as MIMIC-IV and MIMIC-IV-ECG require completion of the official credentialing and data-use process before use.

References:

1. World Health Organization. Cardiovascular diseases (CVDs): key facts. WHO, 2025. Available at: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

2. Virani, S. S., et al. Heart disease and stroke statistics - 2025 update: a report from the American Heart Association. Circulation, 2025. doi:10.1161/CIR.0000000000001303.

3. Goff, D. C., et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk. Journal of the American College of Cardiology, 63(25), 2935-2959, 2014.

4. SCORE2 Working Group. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. European Heart Journal, 42(25), 2439-2454, 2021.

5. Hippisley -Cox, J., Coupland, C., and Brindle, P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease. BMJ, 357, j2099, 2017.

6. Collins, G. S., et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ, 385, e078378, 2024. doi:10.1136/bmj-2023-078378.

7. Moons, K. G. M., et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ, 388, e082505, 2025. doi:10.1136/bmj-2024-082505.

8. Van Calster, B., et al. Calibration: the Achilles heel of predictive analytics. BMC Medicine, 17, 230, 2019.

9. Vickers, A. J. and Elkin, E. B. Decision curve analysis: a novel method for evaluating prediction models. Medical Decision Making, 26(6), 565-574, 2006.

10. Little, R. J. A. and Rubin, D. B. Statistical Analysis with Missing Data. 3rd ed. Wiley, 2019.

11. Steyerberg, E. W. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. 2nd ed. Springer, 2019.

12. Wagner, P., et al. PTB-XL, a large publicly available electrocardiography dataset. Scientific Data, 7, 154, 2020. doi:10.1038/s41597-020-0495-6.

13. Johnson, A. E. W., et al. MIMIC-IV, a freely accessible electronic health record dataset. Scientific Data, 10, 1, 2023. doi:10.1038/s41597-022-01899-x.

14. Centers for Disease Control and Prevention. 2024 BRFSS Survey Data and Documentation. CDC, 2025.

15. Centers for Disease Control and Prevention, National Center for Health Statistics. National Health and Nutrition Examination Survey data and documentation. CDC/NCHS, 2024.

16. Framingham Heart Study. Available FHS data. Boston University and National Heart, Lung, and Blood Institute.

УДК 004.414.23:004.855.5

Жиенбай Аргын

Магистрант
Научный руководитель: Кашкимбаева Н.Ж.
PhD, доцент
кафедра компьютерной инженерии
Astana IT University,
(г. Астана, Казахстан)

ВЛИЯНИЕ СТРАТЕГИЙ КОНСТРУИРОВАНИЯ ПРОМПТОВ НА КАЧЕСТВО АВТОМАТИЗИРОВАННОГО РЕВЬЮ КОДА: СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПОДХОДОВ ZERO-SHOT, FEW-SHOT И CHAIN-OF-THOUGHT

Аннотация. В работе исследуется влияние стратегий конструирования промптов на качество автоматизированного ревью pull request на основе больших языковых моделей. Проведён контролируемый эксперимент с тремя условиями — zero-shot, few-shot (два примера) и chain-of-thought (CoT) — на 100 реальных pull request из девяти Python-репозиториев. Применялись метрики BERTScore F1 и ROUGE-L, а также парный t-тест и d Коэна. Установлено, что CoT достигает наилучших результатов (ROUGE-L = 0,0298; BERTScore = 0,8241) и значимо превосходит zero-shot ($p < 0{,}001$; $d = 0{,}543$). Выявлен эффект краткости: CoT генерирует наиболее лаконичные ответы (42,6 слова), близкие к человеческому эталону (30,5 слова).

Ключевые слова: промпт-инжиниринг, ревью кода, большие языковые модели, chain-of-thought, few-shot, ROUGE-L.

Введение. Автоматизированное ревью кода на основе больших языковых моделей (LLM) является активно развивающимся направлением в программной инженерии [1]. Модели, такие как CodeLlama [2] и DeepSeekCoder [3], способны генерировать комментарии к изменениям кода в режиме нулевого промпта, однако систематически упускают ошибки логики и производительности [4]. Одним из наиболее доступных способов повышения качества генерации, не требующим дообучения или внешних инструментов, является выбор стратегии конструирования промпта. Chain-of-thought (CoT) промптинг, введённый Wei et al. [5], и few-shot обучение в контексте, предложенное Brown et al. [6], показали значительные улучшения на задачах рассуждения и генерации. Тем не менее их сравнительная эффективность для задачи ревью pull request количественно не исследовалась. Настоящая работа восполняет данный пробел.

Материалы и методы. Датасет сформирован из 155 реальных pull request, собранных через GitHub REST API v3 из девяти открытых Python-репозиториев: psf/requests, pallets/flask, django/django, scikit-learn/scikit-learn, pandas-dev/pandas, tiangolo/fastapi, pydantic/pydantic, sqlalchemy/sqlalchemy, celery/celery [7]. Каждая запись содержит унифицированный diff (усечённый до 3 000 символов) и соответствующий inline-комментарий человека-рецензента. Две записи выделены как held-out примеры для few-shot условия. Из оставшихся 153 записей случайно выбраны 100 (seed = 42). Во всех

условиях использовалась одна модель — claude-haiku-4-5 (Anthropic), что позволяет изолировать эффект формулировки промпта от модельных различий.

Оценивались три условия: (1) Zero-Shot — только diff с инструкцией «напиши краткое ревью в 1–3 предложениях»; (2) Few-Shot — два примера (diff → комментарий человека) перед целевым запросом; (3) Chain-of-Thought — трёхшаговый промпт: Шаг 1 — описать изменения, Шаг 2 — выявить проблемы, Шаг 3 — написать финальный комментарий (только текст после «Шаг 3:» использовался как гипотеза).

Метрики: BERTScore F1 (модель roberta-large) [8] и ROUGE-L F1 [9]. Статистическая значимость проверялась парным t-тестом ($\alpha = 0{,}05$); размер эффекта — по d Коэна [10]: $d < 0{,}2$ — незначимый; 0,2–0,5 — малый; 0,5–0,8 — средний.

Результаты. В таблице 1 представлены значения ROUGE-L и BERTScore F1 по всем условиям. CoT достигает наилучших показателей по обеим метрикам (ROUGE-L = 0,0298; BERTScore = 0,8241), за ним следует Few-Shot (ROUGE-L = 0,0247; BERTScore = 0,8203) и Zero-Shot (ROUGE-L = 0,0213; BERTScore = 0,8166). Порядок CoT > Few-Shot > Zero-Shot устойчив для обеих метрик.

Таблица 1 – Результаты оценки по условиям промптинга (n = 100)

Условие	ROUGE-L	Станд. откл.	BERTScore F1
Zero-Shot	0,0213	0,0113	0,8166
Few-Shot	0,0247	0,0146	0,8203
Chain-of-Thought	0,0298	0,0189	0,8241

Таблица 2 содержит результаты парных t-тестов. Установлено, что CoT значимо превосходит Zero-Shot по ROUGE-L со средним эффектом ($t = -5{,}43$; $p < 0{,}001$; $d = 0{,}543$). Few-Shot также значимо превосходит Zero-Shot, но с малым эффектом ($t = -2{,}52$; $p = 0{,}013$; $d = 0{,}252$). Различие между CoT и Few-Shot значимо ($t = -3{,}06$; $p = 0{,}003$; $d = 0{,}306$).

Таблица 2 – Результаты парных t-тестов и d Коэна (n = 100, α = 0,05)

Сравнение	Метрика	t	p	d	Значимо?
Zero-Shot vs. Few-Shot	ROUGE-L	−2,52	0,013	0,252 (малый)	Да
Zero-Shot vs. CoT	ROUGE-L	−5,43	<0,001	0,543 (средний)	Да
Few-Shot vs. CoT	ROUGE-L	−3,06	0,003	0,306 (малый)	Да
Zero-Shot vs. CoT	BERTScore	−5,58	<0,001	0,558 (средний)	Да

Выявлен эффект краткости: CoT генерирует наиболее короткие финальные комментарии (42,6 слова в среднем), ближайшие к человеческому эталону (30,5 слова), тогда как Zero-Shot производит наиболее длинные ответы (88,3 слова). Few-Shot занимает промежуточное положение (69,3 слова). Коэффициенты многословности относительно эталона: Zero-Shot — 2,9×, Few-Shot — 2,3×, CoT — 1,4×. Трёхшаговая структура CoT функционирует как неявное ограничение длины: финальный шаг («напишите один лаконичный комментарий») направляет модель к краткому целевому выводу, что повышает долю совпадающих токенов с кратким человеческим эталоном и тем самым улучшает ROUGE-L [7].

Обсуждение. Улучшение CoT по ROUGE-L (d = 0,543) складывается из двух компонентов. Первый — структурный: декомпозиция на шаги принуждает модель сначала явно определить суть изменений, а затем сформулировать конкретный вывод, снижая долю обобщённых фраз. Второй — обусловленный длиной: более краткий гипотезный текст при одинаковой смысловой точности получает более высокий ROUGE-L вследствие его чувствительности к длине гипотезы через компонент точности. Разделить эти компоненты в рамках данного эксперимента невозможно без контрольного условия с явным ограничением длины.

Few-Shot улучшает Zero-Shot за счёт стилистического ограничения: два демонстрационных примера с лаконичными комментариями индуцируют модель к более короткому выводу. Эффект меньше, чем у CoT, поскольку сигнал двух примеров слабее, чем явная структурная декомпозиция. Применение ретривально-отобранных примеров, семантически близких к текущему diff, ожидаемо даст более значительный прирост [11].

С практической точки зрения CoT рекомендуется как стратегия промптинга по умолчанию в системах автоматизированного ревью кода: она не требует дополнительных данных, внешних инструментов или дообучения, обеспечивает статистически значимое улучшение со средним эффектом и неявно снижает избыточную многословность, устраняя конфаунд, ранее зафиксированный в оценке генерации ревью [7].

Заключение.

1. Установлено, что Few-Shot промптинг с двумя примерами значимо превосходит Zero-Shot по ROUGE-L (p = 0,013; d = 0,252, малый эффект). Улучшение объясняется стилистическим ограничением, индуцируемым демонстрационными примерами.

2. Chain-of-Thought достигает наилучших показателей среди всех условий (ROUGE-L = 0,0298; BERTScore = 0,8241) и значимо превосходит Zero-Shot (p < 0,001; d = 0,543, средний эффект) и Few-Shot (p = 0,003; d = 0,306, малый эффект).

3. Выявлен эффект краткости: CoT генерирует наиболее лаконичные ответы (42,6 слова; 1,4× от эталона), тогда как Zero-Shot — наиболее объёмные (88,3 слова; 2,9×). Структурная декомпозиция CoT функционирует как неявное ограничение длины, улучшающее ROUGE-L и снижающее конфаунд многословности.

4. CoT рекомендуется как стратегия промптинга по умолчанию для LLM-систем ревью кода: она обеспечивает улучшение качества при нулевых дополнительных затратах.

Список использованной литературы:

1. Tufano M. et al. Using pre-trained models to boost code review automation // Proceedings of ICSE. – 2022. – P. 1–12.

2. Rozière B. et al. Code Llama: Open foundation models for code // arXiv:2308.12950. – 2023.

3. Guo D. et al. DeepSeek-Coder: When the large language model meets programming // arXiv:2401.14196. – 2024.

4. Zhienbay A., Kashkimbayeva N. Automated code review using large language models: A zero-shot comparative evaluation // Astana IT University Technical Report. – 2024.

5. Wei J. et al. Chain-of-thought prompting elicits reasoning in large language models // Proceedings of NeurIPS. – 2022. – Vol. 35. – P. 24824–24837.

6. Brown T. et al. Language models are few-shot learners // Proceedings of NeurIPS. – 2020. – Vol. 33. – P. 1877–1901.

7. Zhienbay A., Kashkimbayeva N. Evaluating hybrid approaches to automated code review: Static analysis and RAG as contextual enrichment for LLMs // Journal of Astana IT University. – 2025.

8. Zhang T. et al. BERTScore: Evaluating text generation with BERT // Proceedings of ICLR. – 2020.

9. Lin C.-Y. ROUGE: A package for automatic evaluation of summaries // Proceedings of ACL Workshop. – 2004. – P. 74–81.

10. Cohen J. Statistical Power Analysis for the Behavioral Sciences. – 2nd ed. – Hillsdale: Lawrence Erlbaum Associates, 1988. – 567 p.

11. Nashid N., Sintaha M., Mesbah A. Retrieval-based prompt selection for code-related few-shot learning // Proceedings of ICSE. – 2023. – P. 2450–2462.

UDC 004.4

Orazgali Akhan Askarovich,

Iskakova Kamila Yermekovna,

Pernebay Aiym Yerzhankyzy

Students
Scientific Supervisor: Orazova Arailym Zh.
Senior Lecturer, MSc in Technical Sciences
School of Software Engineering
Astana IT University
(Astana, Kazakhstan)

DEVELOPMENT OF A CROSS-PLATFORM MOBILE APPLICATION FOR AUTOMATED TRACKING AND MONITORING OF FOOD EXPIRATION DATES USING BARCODE SCANNING AND NOTIFICATION TECHNOLOGIES

Abstract: Food waste represents a critical global challenge, with households consistently identified as a primary source of avoidable waste due to insufficient monitoring of product expiration dates. This paper presents the design and implementation of Food Tracker — a cross-platform mobile application for automated tracking and monitoring of food expiration dates. The application is built on a modular monolithic architecture using Flutter, Python FastAPI, MongoDB, and Docker, and integrates barcode scanning via the mobile_scanner library, product metadata retrieval from the OpenFoodFacts database, server-side push notification scheduling through Firebase Cloud Messaging, and AI-powered recipe recommendations delivered by the Spoonacular API with ingredient normalisation performed by the Google Gemini API. A pre-development survey of 63 respondents confirmed strong user demand, with 90.5% reporting occasional food waste and 87.3% expressing that expiry reminders would reduce it. The system achieves over 80% code reuse across Android and iOS platforms, passes a suite of 101 automated backend tests, and is deployed on a cloud PaaS environment. The proposed solution addresses functional gaps identified in existing applications — namely the absence of cross-platform support combined with automated barcode-based product registration and server-side notification scheduling — and offers a reusable architectural template for similar sustainability-oriented mobile systems.

Keywords: food expiration tracking, barcode scanning, push notifications, Firebase Cloud Messaging, Flutter, cross-platform mobile application, food waste reduction.

Introduction. Food waste is one of the most pressing sustainability challenges of the twenty-first century. According to the United Nations Environment Programme, food loss and

waste account for 8–10% of annual global greenhouse gas emissions, while approximately 1.05 billion tonnes of food are wasted globally each year, even as 783 million people face food insecurity [1]. At the household level, this waste is predominantly driven by behavioural factors: consumers purchase quantities that exceed short-term needs, forget items stored in the refrigerator or freezer, and lack efficient mechanisms for tracking the expiration dates of products they have already purchased [2]. These patterns align with findings from behavioural economics research, which links household food waste to present bias and bounded rationality [3].

The widespread adoption of smartphones has created a practical opportunity to address these behavioural patterns through automated digital tools. However, as demonstrated by a review of existing consumer-facing food management applications, no currently available solution simultaneously provides automated barcode-based product registration, cross-platform support for both Android and iOS, server-side push notification scheduling, and integrated recipe suggestions based on inventory content. Applications such as FreshKeep and Foodiry address portions of this requirement set but remain restricted to iOS and rely on manual expiration date entry [4]. Smart kitchen hardware solutions offer broader automation but are inaccessible to the student and home user demographic due to cost [5].

This paper presents Food Tracker, a cross-platform mobile application developed to address these gaps. The application enables users to register food products through barcode scanning or manual entry, track expiration dates, receive user-configurable push notifications before items expire, and access recipe recommendations derived from their current inventory. The system targets students and independent households — a demographic our pre-development survey identified as both highly susceptible to food waste and highly receptive to a dedicated tracking application.

Related Work. Academic and industry literature has identified mobile technology as a promising vector for household food waste reduction. A systematic review published in Frontiers in Digital Health (2024) found that smartphone-based applications consistently improved healthy eating behaviours in randomised controlled trials, particularly where applications incorporated real-time feedback and goal-setting [6]. Research using a micro-randomised trial design further demonstrated that push notifications have measurable effects on user behaviour change, with notification timing and personalisation significantly moderating the effect size [7].

Cross-platform mobile development frameworks have matured substantially over the past decade. Comparative analyses of Flutter, React Native, and Xamarin consistently identify Flutter as the leading option for projects requiring strong visual consistency and near-native rendering performance, owing to its Dart-based native compilation model [8]. Flutter's GitHub repository had attracted over 154,000 stars by 2023, substantially ahead of React Native's 110,000 [9], reflecting its growing adoption among professional developers.

Barcode scanning for consumer product identification relies primarily on the ZXing and Google ML Kit decoding engines. The mobile_scanner Flutter package wraps these engines to provide a unified cross-platform barcode scanning interface [10]. Product metadata resolution is addressed by the OpenFoodFacts open database, a community-maintained repository covering more than 1.7 million products from 150 countries [11]. Image-based product recognition — proposed as an alternative to barcode scanning in several research prototypes — continues to face generalisation challenges under real-world conditions and was not considered for production deployment in this project [12].

A comparative analysis of existing food management applications — evaluated against the criteria of barcode scanning capability, expiration date tracking automation, cross-platform support, notification scheduling, and sustainability orientation — reveals that no currently available solution satisfies all five criteria. Table 1 summarises this comparison.

Table 1 – Comparative Analysis of Existing Food Management Solutions

Feature / Criterion	FreshKeep	Foodiry	Smart Fridge Systems	Food Tracker (proposed)
Barcode scanning	✗	✓	—	✓
Expiry date tracking	Manual	Manual	Partial	Automatic
Cross-platform (Android + iOS)	✗ (iOS only)	✗ (iOS only)	Hardware	✓
Push notifications	Basic	✗	✗	✓ (configurable)
Recipe suggestions	✗	✗	Limited	✓
Sustainability focus	✗	✗	✗	✓
Accessibility (cost)	Free	Free / paid	Very expensive	Free

Research Methodology. The development process followed an iterative and incremental methodology, in which the system was built module by module according to the logical dependency ordering of its components. User authentication was developed first, as all other modules depend on its functionality; the recipe recommendation module was implemented last, as it depends on the output of most other modules. This ordering reduced integration-time rework and provided each subsequent module with a stable tested foundation.

Requirements were elicited through three complementary methods. First, a questionnaire survey of 63 respondents was conducted to characterise the target user demographic and validate the problem framing. The survey population skewed toward younger adults, with 50.8% aged 25–34 and 20.6% aged 18–24; students comprised 54% of respondents. Key findings confirmed the relevance of the problem: 90.5% of participants reported occasionally throwing away food at home, 54% indicated that they forget about food stored in the refrigerator or freezer, and 49.2% reported purchasing duplicate items due to forgetfulness. Demand for the proposed solution was strong: 79.4% expressed willingness to use a food tracking application, 87.3%

agreed that expiry reminders would reduce their food waste, and 84.1% ranked barcode scanning as their most preferred input method.

Second, a structured review of existing food management applications and supporting academic literature was conducted to identify functional gaps in the current application landscape and to ground technology selection decisions in empirical evidence. Third, functional and non-functional requirements were derived from the survey results and the application review, and validated iteratively throughout the development cycle through developer testing and peer feedback sessions.

System Design and Architecture. The backend is implemented as a modular monolith — a single deployable application internally divided into loosely coupled modules. This architectural style was chosen over a microservice architecture for three reasons specific to the project context: the development team is small (three members), the expected initial user base is approximately 1,000 users, and the monolithic deployment model substantially reduces operational overhead. The modular structure ensures that future migration to a microservice architecture remains feasible if the system grows beyond the current scale.

The backend modules and their responsibilities are as follows. The Auth module manages user registration, login, JWT token generation, FCM token storage, and notification preference configuration. The Products module resolves barcodes using a database-first strategy with an OpenFoodFacts fallback, and supports manual product creation. The Inventory module handles CRUD operations on user food inventories, including soft deletion and expiry statistics aggregation. The Notifications module runs a background scheduler that polls the database at 15-minute intervals using a MongoDB \$elemMatch query, identifies items that have reached a notification threshold, and dispatches push notifications via the Firebase SDK. The Recipes module normalises ingredient names using the Gemini API, queries the Spoonacular API, caches results in MongoDB, and ranks recommendations by a match score that prioritises ingredients approaching expiry. The Storage module provides AI-assisted recommended shelf-life durations for products added without a barcode. The Shopping List module provides basic CRUD operations for a user shopping list populated either manually or from recipe detail screens.

The mobile frontend is implemented in Flutter using the Riverpod state management library for reactive, compile-safe provider injection. Navigation is handled by GoRouter with JWT-based redirect logic; network communication uses the Dio HTTP client with request interceptors that inject the stored JWT into every outgoing request and handle token expiry by triggering automatic logout. The application shares over 80% of its codebase between the Android and iOS targets.

The full technology stack is summarised in Table 2.

Table 2 – Technology Stack of the Food Tracker Application

Layer	Technology	Rationale
Mobile frontend	Flutter (Dart)	Single codebase; >80% code reuse; native ARM compilation
Backend framework	Python FastAPI	Async I/O; auto-generated Swagger docs; Pydantic validation
Database	MongoDB	Flexible document model; async Motor driver; JSON-native
Push notifications	Firebase Cloud Messaging	Cross-platform; unified API for Android & iOS
Product metadata	OpenFoodFacts API	1.7M+ products; free; Open Database Licence
Recipe data	Spoonacular API	Ingredient-based search; match scoring
AI text normalisation	Google Gemini API	Converts raw product names to standardised ingredient terms
Containerisation	Docker	Reproducible build; multi-stage image; cloud PaaS deployment

Implementation. The backend is structured as a Python package under a modules/ directory, with each module containing its own router.py, service.py, models.py, and schemas.py components. This separation of concerns simplifies debugging and enables independent unit testing of each layer. The FastAPI application mounts each module's router at a versioned prefix (e.g., /api/v1/inventory/) and applies a global CORS middleware. Authentication and authorisation are enforced through a get_current_user() dependency injected into all protected endpoints, which validates the JWT and returns the authenticated user context.

Product information retrieval follows a three-level strategy. When a barcode is scanned, the backend first checks the MongoDB products collection. On a cache miss, it queries the OpenFoodFacts REST API and caches the response tagged as is_verified: true. If OpenFoodFacts returns no result, a 404 Not Found response is returned to the mobile client, which then presents a manual entry form to the user. This caching strategy eliminates redundant external API calls for frequently scanned products and complies with the OpenFoodFacts terms of use.

The notification scheduler is implemented using APScheduler's BackgroundScheduler, which runs as a non-blocking background thread within the FastAPI process. Every 15 minutes, the scheduler calls a service function that queries the inventory_items collection using a \$elemMatch operator on the scheduled_notifications array, isolating documents where at least one notification entry has a send_at timestamp in the past and a sent flag equal to false. For each matching item, the scheduler retrieves the owner user's FCM token and dispatches a notification payload via the Firebase Admin SDK, then marks the sent flag as true to prevent duplicate delivery. This approach decouples notification logic from the API request path and scales efficiently for inventories of up to several thousand items.

The recipe recommendation pipeline operates in three stages. First, raw product names from the user's active inventory — which may be in multiple languages and contain non-standard formatting — are passed to the Google Gemini API for normalisation into simple English ingredient terms (e.g., "Milk 3.2%" → "milk"). Normalisation results are cached in the ingredient_normalizations collection. Second, the normalised ingredient list is submitted to the Spoonacular API, which returns a ranked list of recipes matching the available ingredients. Third, the backend computes a match_score for each result, adjusted to prioritise recipes whose required ingredients have the earliest expiration dates in the user's inventory. Recipes with equal scores are shuffled daily to introduce variety.

The database design uses MongoDB Atlas with seven collections: users, products, inventory_items, recipes, recipe_queries, ingredient_normalizations, storage_recommendations, and shopping_list_items. Compound indexes — notably (user_id, expiration_date) on inventory_items and (status, scheduled_notifications.send_at) for the notification scheduler — are created at application startup to support the primary access patterns without full collection scans. The backend is containerised using a Docker multi-stage build and deployed to the Render cloud PaaS, with MongoDB Atlas as the managed database service.

Results and Discussion. The fully implemented Food Tracker application provides the following core features: barcode-triggered product registration with automatic metadata retrieval; manual product entry with AI-assisted expiry date suggestion; a personal food inventory with colour-coded expiry status indicators (active, expiring soon, expired); configurable push notifications at thresholds of 1, 2, 3, 5, 7, or 14 days before expiration; recipe recommendations ranked by ingredient match and expiry proximity; a shopping list populable from recipe ingredient screens; and a settings screen with full English/Russian localisation.

The automated backend test suite comprises 101 tests across six test modules — integration tests for the auth, inventory, notifications, and products modules, and unit tests for the security and data models components — all of which pass in 27.06 seconds using the mongomock-motor in-memory database library and mock external API responses. This testing approach achieves high coverage without requiring live external service connections and provides a replicable pattern for FastAPI applications with complex external service dependencies.

The application satisfies the six functional requirements defined during the design phase: inventory CRUD operations (FR-1), barcode-based product registration (FR-2), manual expiration date entry and editing (FR-3), automated pre-expiry notifications (FR-4), structured inventory display (FR-5), and recipe suggestions based on available inventory (FR-6). Non-functional requirements are met as follows: barcode recognition completes within 2 seconds on current mid-range devices using the ML Kit decoding engine (NFR-2); the modular backend structure supports independent extension of each module (NFR-4); the shared Flutter codebase runs on both Android and iOS with over 80% code reuse (NFR-5); and JWT-based

authentication with bcrypt password hashing and HTTPS-only communication satisfies baseline security requirements (NFR-6).

The key limitation of the current implementation is its single-user model, which does not support household sharing. Survey feedback and user acceptance testing identified shared inventory functionality as the most requested future enhancement. Product metadata coverage is also subject to the completeness of the OpenFoodFacts database for a given product category and geographic market, requiring the manual entry fallback for unrecognised barcodes. Notification delivery depends on the availability of Firebase infrastructure (Google Play Services on Android, APNS on iOS), which may not be available on all device configurations.

Conclusion. This paper presented the design, implementation, and evaluation of Food Tracker, a cross-platform mobile application for automated food expiration tracking. The system integrates barcode scanning, OpenFoodFacts product retrieval, APScheduler-driven push notification dispatch via Firebase Cloud Messaging, and Spoonacular-powered recipe recommendations with Gemini-based ingredient normalisation, within a modular monolithic FastAPI backend and a shared Flutter frontend targeting Android and iOS.

A pre-development survey of 63 respondents confirmed strong demand for an automated food tracking solution among students and young independent adults. The implemented system passes 101 automated backend tests, achieves over 80% cross-platform code reuse, and addresses functional gaps — primarily the combination of automated barcode-based inventory management with server-side notification scheduling — that remain unmet by existing consumer applications.

Future development directions identified through the project include the addition of shared household inventory functionality, integration of optical character recognition for automated expiry date extraction from product packaging, and expansion of the analytics module to support trend analysis and predictive food waste insights. The architectural pattern demonstrated in this work — modular monolith backend with APScheduler notifications, MongoDB multi-level caching, and Flutter cross-platform frontend — offers a reusable template for future sustainability-oriented mobile system development.

References:

1. UNEP. Food Waste Index Report 2024. United Nations Environment Programme, Nairobi, 2024.

2. Quested T., Ingle R., Bhawan A. Household food and drink waste in the United Kingdom 2012. WRAP Technical Report, 2013.

3. Danzer A. M., Zeidler H. Food waste and dynamic inconsistency: A behavioral economics perspective. CESifo Working Paper Series, No. 11781, 2025.

4. Ferrara G., Kim J., Lin S., Hua J., Seto E. A focused review of smartphone diet-tracking apps: Usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. JMIR mHealth and uHealth, 7(5): e9232, 2019.

5. Dai X. Robust deep-learning based refrigerator food recognition. Frontiers in Artificial Intelligence, 2024.

6. Gebrie M. H. et al. The use of internet-based smartphone apps consistently improved consumers' healthy eating behaviors: a systematic review of randomized controlled trials. Frontiers in Digital Health, 6: 1282570, 2024.

7. Bell I. H. et al. How notifications affect engagement with a behavior change app: results from a micro-randomised trial. JMIR Mental Health, 2023.

8. Sanders M. Flutter vs React Native vs Xamarin. Medium, 2025. https://medium.com/@may_sanders/flutter-vs-react-native-vs-xamarin-11f7dae64c98

9. Mobile Reality. Flutter vs React Native vs Xamarin. https://themobilereality.com/blog/xamarin-vs-flutter-vs-react-native, 2023.

10. Scanbot SDK. ML Kit vs. ZXing – Comparing two popular barcode scanning libraries. https://scanbot.io/blog/ml-kit-vs-zxing/, 2025.

11. Open Food Facts. Data, API and SDKs. https://world.openfoodfacts.org/data, 2024.

12. Bu Y. et al. Recognition of food images based on transfer learning and ensemble learning. PLOS ONE, 19(1): e0296789, 2024.

UDC 004.85:338.27

Tussupov Yersain

9th grade student
PhysTech School
(Almaty, Kazakhstan)
Scientific Supervisor: Toksanbayev Bauyrzhan
Master's student
Computer Science
Nazarbayev University
(Astana, Kazakhstan)

THE USE OF NEURAL NETWORKS FOR FORECASTING CRYPTOCURRENCY PRICES

Abstract. This article examines the application of LSTM-type recurrent neural networks for forecasting cryptocurrency price dynamics. It describes the model architecture, the factors influencing digital asset prices, and the results of testing on Bitcoin data for 2018–2023.

Keywords: neural networks, LSTM, cryptocurrency, Bitcoin, forecasting, machine learning.

1. Introduction

Cryptocurrencies are a new class of digital assets based on blockchain technology. Since the emergence of Bitcoin in 2009, the digital currency market has grown to more than twenty thousand coins with a combined market capitalization exceeding one trillion dollars. The high volatility of this market makes price forecasting a task of particular interest to both investors and researchers.

Classical methods of financial analysis show limited effectiveness when applied to cryptocurrencies because of their specific nature: round-the-clock trading, the absence of a regulator, and a strong dependence on the information climate in social media. Deep learning methods — recurrent neural networks above all — open up new possibilities here [1]. The aim of this work is to investigate the applicability of the LSTM architecture for forecasting the Bitcoin exchange rate and to assess its accuracy in comparison with traditional approaches.

2. Neural Networks and the LSTM Architecture

A neural network is a mathematical model built by analogy with biological neurons. It consists of an input layer, hidden layers, and an output layer, connected by weighting coefficients that are tuned during training through the backpropagation method. Standard neural networks are poorly suited to working with time series, since they do not take into account the order in which data arrives.

A recurrent network with long short-term memory (LSTM) solves this problem through a special cell-state mechanism and three control gates. The forget gate discards unnecessary information from previous steps, the input gate writes new data to memory, and the output gate

determines what is passed on to the next step. Thanks to this, LSTM captures dependencies over long time intervals — precisely what is needed for analyzing price series [2, p. 1735].

3. Methodology and Model Architecture

The input data consisted of daily Bitcoin quotes (opening price, closing price, high, low, and trading volume) from January 2018 to December 2023, obtained through the public API of the Binance exchange. Additional technical indicators were calculated: 7-day and 30-day moving averages, the RSI (14 periods), and the MACD. All features were normalized to the [0; 1] range using the min-max method.

The model takes as input a 60-day time window (10 features). The first layer consists of 128 LSTM cells with a dropout of 0.3 to prevent overfitting. The second layer consists of 64 LSTM cells to extract more abstract dependencies. The output fully connected layer produces a price forecast for 1 or 7 days ahead. Training was carried out using the Adam optimizer, with MSE as the loss function, 100 epochs, and a batch size of 32. The test set accounted for 20% of the data.

4. Results

A comparison of the developed model with alternative methods is presented in Table 1.

Table 1 — Comparison of Forecasting Methods

Method	RMSE (USD)	R^2	Trend Accuracy
ARIMA	4,213	0.741	55.1%
Random Forest	2,956	0.862	61.8%
Single-layer LSTM	2,301	0.891	63.9%
Two-layer LSTM (proposed model)	1,847	0.913	67.4%

The two-layer LSTM model achieved an RMSE more than twice as low as that of ARIMA and improved trend-direction forecasting accuracy by 12.3 percentage points. The main limitations of the model are its inability to account for sudden “black swan” events and a decline in accuracy at forecast horizons beyond seven days.

5. Conclusion

This study has shown that a two-layer LSTM network effectively identifies nonlinear dependencies in cryptocurrency price series and significantly outperforms classical statistical methods. At the same time, a neural network is not a universal tool: it should be regarded as an analytical aid rather than a substitute for sound investment decisions. Prospects for further research include integrating social media data and applying transformer architectures.

References:

1. Goodfellow I., Bengio Y., Courville A. Deep Learning. — Cambridge: MIT Press, 2016. — 800 p.

2. Hochreiter S., Schmidhuber J. Long Short-Term Memory // Neural Computation. — 1997. — Vol. 9, № 8. — P. 1735–1780.

UDC 004.8

Seitkadyrov Amirlan
High School Student
(Astana, Kazakhstan)
Zhou Chenliang
Postdoctoral Researcher
(Cambridge, England)

IMPACT OF ANCHORED PROMPTS ON SUMMARIZATION QUALITY ACROSS LARGE LANGUAGE MODELS

Abstract: Anchored prompting-augmenting a summarization prompt with keyphrases extracted from the source-has been reported to improve faithfulness and salience in prompt-based summarization. However, its robustness across model scales and domains remains unclear. We present a controlled evaluation of anchored prompts across five language models spanning 88M to 671B parameters (DistilGPT-2, GPT-2 Medium, BART-base, BART-Large-CNN, and DeepSeek-V3) on two domains: news (CNN/DailyMail) and scientific writing (arXiv). For each dataset, we test 50 long documents and construct anchors using 3-7 KeyBERT keyphrases per document, comparing a minimal baseline prompt against an anchored prompt template. Across both datasets, anchored prompts are not reliably beneficial. Improvements for stronger summarization-tuned models are small and inconsistent, while BART-base exhibits large degradations. Paired bootstrap confidence intervals and Wilcoxon signed-rank tests on document-level ROUGE show that anchored prompting significantly reduces ROUGE for BART-base on both datasets (e.g., ΔROUGE-L $\approx$ -0.026 on CNN/DailyMail and -0.020 on arXiv), and DeepSeek-V3 shows a negative mean shift on arXiv, although this change is not statistically significant in our sample. Qualitative inspection reveals recurrent failure modes: redundant or low-utility keyphrases (often entity-heavy), prompt-induced verbosity, repetition, and topic drift. These results suggest that anchor prompting should be paired with stronger keyphrase extraction and lighter prompt scaffolding, and should be reported with uncertainty estimates rather than single average scores.

Keywords: anchored prompting, prompt engineering, abstractive summarization, keyphrase extraction, KeyBERT, ROUGE, CNN/DailyMail, arXiv, BART, GPT-2, DeepSeek-V3, statistical significance testing

1. Introduction

Automatic text summarization is a core NLP task that compresses long documents into concise summaries while preserving key information [1]. With the widespread use of large language models (LLMs), prompt design has become a practical lever for improving output quality without parameter updates [2, 3]. One widely used idea is anchored prompting: injecting keyphrases from the source into the prompt to steer the model toward salient content. Prior work reports gains from salient/keyword-based prompting [4], but it is unclear whether such gains are robust across (i) model scale and training regime (summarization-tuned vs. general LMs) and (ii) domain (news vs. scientific writing).

This paper provides a controlled comparison of anchored prompting across five models of substantially different capacity and training: GPT-2 Medium, DistilGPT-2, BART-base, BART-Large-CNN, and DeepSeek-V3. We evaluate on two datasets with different discourse

structure and target summaries: CNN/DailyMail (news) [5] and arXiv (scientific abstracts) [6]. Anchors are constructed using KeyBERT [7] with 3-7 keyphrases per document, allowing us to isolate the effect of adding extracted anchors to an otherwise fixed prompting template.

Our main contributions are:

• **Cross-scale, cross-domain evaluation:** anchored prompts are tested on five models ranging from 88M to 671B parameters on both news and scientific summarization.

• **Uncertainty-aware reporting:** beyond average ROUGE, we report paired bootstrap confidence intervals and Wilcoxon signed-rank significance tests for prompt-induced performance differences.

• **Qualitative error analysis:** we provide concrete examples showing how anchors can help (better topical focus) or hurt (repetition, verbosity, and drift), especially when extracted keyphrases are redundant or entity-heavy.

Contrary to the expectation that larger or summarization-tuned models would consistently benefit from anchored prompts, we find no uniform improvement and observe substantial degradations for some settings (notably BART-base). These findings highlight that anchored prompting is not a plug-and-play technique: its effectiveness depends on both the quality of the extracted anchors and the model's ability to use additional prompt structure.

The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 describes our methodology, Section 4 reports quantitative results and significance tests, Section 5 presents qualitative analysis and discusses implications and limitations, and Section 6 concludes with directions for future work.

2. Related Work

Prompt engineering boosts LLM performance without changing the model parameters [2, 3]. In summary, [4] introduced Salient Information Prompting, using keyphrases to improve ROUGE scores by focusing on key content, with their Keyphrase Signal Extractor (SigExt) showing consistent gains across datasets and LLMs. Likewise, [6] created a discourse-aware attention model for summarizing long documents, highlighting the role of the structure. [8] studied question-answering prompts to enhance summarization coherence, while [9] explored anchor-based LLMs, noting advantages in structured tasks.

Scaling laws indicate that larger models perform better on complex tasks because of their greater capacity and sample efficiency [10], suggesting that advanced LLMs might better utilize anchored prompts. Our study tests this idea by comparing performance across models of different sizes and training methods, including GPT-2 Medium [11], BART [12], DistilGPT-2 [13], and DeepSeek-V3 [14], and examines why our results diverge from previous studies.

3. Methodology

We tested anchored prompts on two datasets: CNN/DailyMail, a news dataset with human-written summaries [5], and arXiv, a collection of scientific papers with abstracts [6]. We chose a 50-article subset from each (articles over 1500 words) to ensure complexity.

3.1 Models

We evaluated five LLMs with varying complexities.

• **GPT-2 Medium (355M parameters):** We use the publicly released GPT-2 Medium checkpoint; release and deployment considerations are discussed in [11].

• **BART-Large-CNN (406M parameters):** A transformer-based model fine-tuned for summarization [12].

• **BART-base (139M parameters):** A smaller BART variant [12].

• **DistilGPT-2 (82M parameters):** A GPT-2 checkpoint trained via knowledge distillation; we follow the distillation approach described in [13].

• **DeepSeek-V3:** A large-scale model with 671 billion parameters [14].

3.2 Prompt Design

Keyphrases (3-7 per article) were extracted for the anchored prompts. For GPT-2 Medium, BART-Large-CNN, BART-base, DistilGPT-2, and DeepSeek-V3, we applied KeyBERT [7], using BERT embeddings for semantically rich keyphrases. We chose KeyBERT for all models because it leverages BERT embeddings to extract keyphrases that are semantically meaningful and contextually relevant, making it suitable for a variety of LLM architectures.

The baseline prompt was: "Summarize the following article: [article]". The anchored prompt was: "What is the article about? What happened? Where and when? Why is it important? Considering key aspects: [keyphrases]. Now summarize the article: [article]". The input was limited to 1024 tokens, with output summaries capped at 150 tokens.

We note that this anchored template is intentionally question-heavy; future work will test a compressed anchor template to isolate the effect of prompt complexity.

3.3 Evaluation Metrics

We evaluate summaries with standard automatic metrics:

• **ROUGE-1, ROUGE-2, ROUGE-L:** lexical overlap with the reference summaries.

• **BERTScore:** semantic similarity using contextual embeddings.

• **METEOR:** overlap with synonymy and stemming.

3.4 Statistical Significance and Multiple Comparisons

To assess whether anchored prompting produces reliable changes beyond noisy fluctuations, we test prompt effects using paired comparisons on document-level ROUGE. For each model-dataset pair, we compute per-document differences

Δi = ROUGE-Li (anchored) − ROUGE-Li (baseline) (eq1)

over the same 50 documents.

We report (i) the mean Δ, (ii) a paired bootstrap 95% confidence interval for the mean (20,000 resamples with replacement; seed=42), and (iii) a two-sided Wilcoxon signed-rank test p-value, which does not assume normality of Δi. Because we run 10 hypothesis tests (5 models × 2 datasets), we additionally apply the Holm-Bonferroni procedure to control the family-wise error rate at alpha=0.05.

4. Experiments and Results

We compared baseline and anchored prompts across all models on both CNN/DailyMail and arXiv. The results are summarized in Tables 1 and 2.

Table 1 – Summarization Performance on CNN/DailyMail

Model	Prompt	ROUGE-1	ROUGE-2	ROUGE-L	BERT Score	METEOR
GPT-2 Medium	Baseline	0.109	0.049	0.07	0.831	0.215
GPT-2 Medium	Anchored	0.109	0.048	0.071	0.826	0.213
BART-Large-CNN	Baseline	0.359	0.140	0.240	0.865	0.316

BART-Large-CNN	Anchored	0.362	0.138	0.242	0.863	0.319
BART-base	Baseline	0.295	0.103	0.179	0.847	0.330
BART-base	Anchored	0.264	0.077	0.153	0.826	0.281
DistilGPT-2	Baseline	0.108	0.048	0.069	0.831	0.217
DistilGPT-2	Anchored	0.108	0.048	0.07	0.826	0.214
DeepSeek-V3	Baseline	0.304	0.096	0.181	0.846	0.297
DeepSeek-V3	Anchored	0.302	0.092	0.181	0.832	0.260

Table 2 – Summarization Performance on arXiv

Model	**Prompt**	**ROUGE-1**	**ROUGE-2**	**ROUGE-L**	**BERT Score**	**METEOR**
GPT-2 Medium	Baseline	0.251	0.109	0.141	0.827	0.313
GPT-2 Medium	Anchored	0.250	0.107	0.140	0.824	0.311
BART-Large-CNN	Baseline	0.336	0.105	0.194	0.834	0.188
BART-Large-CNN	Anchored	0.323	0.097	0.186	0.832	0.175
BART-base	Baseline	0.332	0.088	0.182	0.826	0.209
BART-base	Anchored	0.304	0.067	0.161	0.818	0.188
DistilGPT-2	Baseline	0.246	0.107	0.138	0.827	0.311
DistilGPT-2	Anchored	0.245	0.105	0.137	0.824	0.308
DeepSeek-V3	Baseline	0.333	0.075	0.172	0.821	0.193
DeepSeek-V3	Anchored	0.303	0.073	0.164	0.813	0.174

Tables 1-2 report average performance for baseline vs. anchored prompting. Across CNN/DailyMail, anchored prompting produces negligible average changes for GPT-2 Medium, DistilGPT-2, BART-Large-CNN, and DeepSeek-V3, while BART-base consistently degrades. On arXiv, anchored prompting again reduces BART-base performance and can also reduce DeepSeek-V3 performance depending on the metric. To avoid over-interpreting small average differences, we report uncertainty-aware tests on document-level ROUGE in Table 3. GPT-2 Medium shows a statistically significant but practically negligible change on CNN/DailyMail (Δ ROUGE-L $\approx +0.0012$).

Table 3 – Significance testing for ROUGE-L differences (Δ = Anchored - Baseline)

Model	Dataset	ΔROUGE-L (95% CI)	p
DeepSeek-V3	CNN/DailyMail	0.0001 [-0.0169, 0.0158]	0.6119
BART-base	CNN/DailyMail	-0.0257 [-0.0388, -0.0127]	0.0002
BART-Large-CNN	CNN/DailyMail	0.0014 [-0.0151, 0.0180]	0.8789
DistilGPT-2	CNN/DailyMail	0.0004 [-0.0005, 0.0013]	0.4964
GPT-2 Medium	CNN/DailyMail	0.0012 [0.0003, 0.0020]	0.0063
DeepSeek-V3	arXiv	-0.0085 [-0.0209, 0.0030]	0.3989
BART-base	arXiv	-0.0203 [-0.0308, -0.0111]	<0.0001
BART-Large-CNN	arXiv	-0.0081 [-0.0203, 0.0040]	0.1302
DistilGPT-2	arXiv	-0.0010 [-0.0031, 0.0009]	0.6051
GPT-2 Medium	arXiv	-0.0009 [-0.0027, 0.0009]	0.4606

Table 3 shows uncertainty-aware prompt effects for ROUGE-L. On CNN/DailyMail, anchored prompting has no reliable impact for DeepSeek-V3, BART-Large-CNN, DistilGPT-2, and the effect sizes are near zero with confidence intervals spanning zero. BART-base exhibits a clear and practically meaningful degradation (ΔROUGE-L = -0.0257, 95% CI [-0.0389, -0.0127], p=0.0002). GPT-2 Medium shows a statistically significant but practically negligible improvement (ΔROUGE-L ≈ +0.0012), which does not remain significant after Holm-Bonferroni correction (adjusted p ≈ 0.051).

On arXiv, BART-base again degrades significantly (ΔROUGE-L = -0.0203, 95% CI [-0.0312, -0.0111], p < 0.0001), and this effect remains significant after Holm-Bonferroni correction. Other models show negative mean shifts whose confidence intervals include zero, indicating no robust evidence that anchoring improves performance in our sample. Overall, these results suggest that anchored prompting is not a consistent upgrade and can substantially harm certain model-domain combinations, especially when the anchor list is redundant or misaligned with the reference summary.

5. Qualitative Analysis

Automatic metrics summarize average behavior but can hide systematic failure modes. We therefore inspect representative documents with large positive/negative ΔROUGE-L to understand when anchors help versus harm. Across models, we observe three recurring patterns: (i) redundant/entity-heavy anchors that bias generation toward repetition, (ii) prompt-induced verbosity that crowds out key facts, and (iii) topic drift when extracted keyphrases are only weakly aligned with the reference summary.

5.1 Case Studies

5.1.1 Case A (CNN/DailyMail, BART-Large-CNN): anchors improve topical focus

• **Reference (excerpt):** The World Happiness Report highlights the happiest countries; the UN declared World Happiness Day in 2012.

• **Keyphrases:** national happiness; world happiness; countries happy; happiness day; recognizing happiness.

• **Baseline (excerpt):** Switzerland took the top spot from Denmark in 2015 ... The U.N. General Assembly declared March 20 as World Happiness Day in 2012.

• **Anchored (excerpt):** People who live in the happiest countries have longer life expectancy ... Switzerland took the top spot from Denmark in 2015...

• **Observation:** The anchored version foregrounds the main frame (happiness report and its implications), matching the reference more directly.

5.1.2 Case B (CNN/DailyMail, BART-Large-CNN): anchors cause detail loss and generic phrasing

• **Reference (excerpt):** Former Australian captain Richie Benaud died in a Sydney hospice; tributes followed.

• **Keyphrases:** cricket commentator; Richie Benaud; former Australia captain; Benaud died.

• **Baseline (excerpt):** The iconic broadcaster passed away peacefully in his sleep ... tributes flowed in...

• **Anchored (excerpt):** Legendary commentator and former Australia captain Richie ... one of the world's most recognised commentators...

• **Observation:** The anchored output shifts toward generic biography-style wording and drops concrete reference-aligned details (timing, hospice context), reducing overlap.

5.1.3 Case C (CNN/DailyMail, BART-base): entity-heavy anchors amplify repetition and truncation

• **Reference (excerpt):** Ronaldo scored five times in Real Madrid's 9-1 win; extreme scoring run.

• **Keyphrases:** Ronaldo scored; Granada Ronaldo; fixtures Ronaldo; regularly Ronaldo.

• **Baseline (excerpt):** Even by his stratospheric standards ... an eight-minute hat-trick ...

• **Anchored (excerpt):** Even by his stratospheric standards ... Ronaldo was back to his brilliant, unplayable best as he scored five...

• **Observation:** For BART-base, anchors provide little new semantic guidance and encourage repetitive surface forms, often shortening or truncating the output and hurting alignment with the reference.

5.1.4 Case D (arXiv, DeepSeek-V3): anchors increase boilerplate and reduce specificity

• **Reference (excerpt):** dynamics of a single-level quantum dot driven out of equilibrium; three relaxation time scales.

• **Keyphrases:** single electrons; quantum electron optics; electrons mesoscopic; nanoelectronics.

• **Baseline (excerpt):** The article investigates relaxation dynamics of a quantum dot driven out of equilibrium ... governed by charge, spin, and an additional time scale ...

• **Anchored (excerpt):** The article explores control and manipulation of single electrons ... (high-level framing and broader context) ...

• **Observation:** Anchors bias the model toward a generic "single-electron control" framing and add boilerplate context, while the baseline more directly mirrors the reference's specific contribution.

Overall, these cases suggest that anchor prompting is sensitive to anchor quality. When keyphrases capture the true discourse spine of the reference (Case A), anchors can help; when anchors are redundant or overly entity-centric (Cases B-C) or too general for technical contributions (Case D), they can harm faithfulness and specificity.

6. Discussion

Our results indicate that anchored prompting with KeyBERT keyphrases is not a universally reliable improvement for summarization. The quantitative analysis shows that most average differences are small and often not statistically distinguishable from zero, while certain settings exhibit clear, significant degradations (notably BART-base on both datasets). The qualitative inspection clarifies why: anchors frequently become redundant (entity-heavy or paraphrastic), increase prompt verbosity, and can induce repetition or topic drift.

6.1 Why do our results differ from prior positive findings?

Prior work on salient prompting reports more consistent gains [4]. A plausible explanation is that anchor construction matters as much as anchor injection. Off-the-shelf keyphrase extraction may not reliably identify the concepts that match the reference summaries, especially in technical abstracts where the key contribution is often a specific method, constraint, or result rather than a topical keyword. In addition, our anchored template includes multiple guiding questions plus the keyphrase list; this extra structure can crowd the model's context window and encourage generic "answer-the-questions" behavior rather than faithful summarization.

6.2 Implications for prompt-based summarization

Anchored prompting should be treated as a hypothesis to be validated per model and domain, not a default upgrade. In practice, anchor prompting may require (i) deduplication and filtering of entity-only anchors, (ii) shorter scaffolding (fewer leading questions), and (iii) domain-adapted keyphrase extraction. Reporting prompt effects with uncertainty estimates is also important: some statistically significant differences we observe are extremely small in magnitude and may not be practically meaningful.

6.3 Limitations and future work

This study uses a 50-document subset per dataset and a single anchor extraction method (KeyBERT) and anchor template. We mainly provide significance testing for ROUGE; future work should extend uncertainty-aware analysis to semantic metrics and include human evaluation (e.g., faithfulness and coverage ratings or pairwise preferences) to better capture real-world quality differences. Further improvements could also test specialized keyphrase extractors, lighter anchor templates, and alternative steering methods such as QA-style prompting [8].

7. Conclusion

We presented a controlled, cross-scale evaluation of anchored prompting for summarization across five models and two domains. Anchored prompts built with KeyBERT do not yield consistent gains and can significantly degrade performance for certain models and datasets, as supported by paired significance tests and qualitative evidence. These findings emphasize that prompt-based steering must be paired with high-quality anchors and uncertainty-aware reporting to avoid overstating small or unstable improvements.

References:

1. Nallapati R., Zhai F., Zhou B. SummaRuNNer: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents [Electronic resource] // arXiv preprint arXiv:1611.04230. — 2016. — URL: https://doi.org/10.48550/arXiv.1611.04230 (accessed: 19.06.2026).

2. Brown T.B., Mann B., Ryder N., Subbiah M., Kaplan J., Dhariwal P. et al. Language Models are Few-Shot Learners [Electronic resource] // arXiv preprint arXiv:2005.14165. — 2020. — URL: https://doi.org/10.48550/arXiv.2005.14165 (accessed: 19.06.2026).

3. Wei J., Wang X., Schuurmans D., Bosma M., Ichter B., Xia F. et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models [Electronic resource] // arXiv preprint arXiv:2201.11903. — 2022. — URL: https://doi.org/10.48550/arXiv.2201.11903 (accessed: 19.06.2026).

4. Xu L., Karim M.A., Dingliwal S., Elangovan A. Salient Information Prompting to Steer Content in Prompt-based Abstractive Summarization [Electronic resource] // Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track. — 2024. — URL: https://doi.org/10.18653/v1/2024.emnlp-industry.4 (accessed: 19.06.2026).

5. Hermann K.M., Kocisky T., Grefenstette E., Espeholt L., Kay W., Suleyman M. et al. Teaching Machines to Read and Comprehend [Electronic resource] // arXiv preprint arXiv:1506.03340. — 2015. — URL: https://doi.org/10.48550/arXiv.1506.03340 (accessed: 19.06.2026).

6. Cohan A., Dernoncourt F., Kim D.S., Bui T., Chang W., Goharian N. A Discourse-Aware Attention Model for Abstractive Summarization of Long Documents [Electronic resource] // arXiv preprint arXiv:1804.05685. — 2018. — URL: https://doi.org/10.48550/arXiv.1804.05685 (accessed: 19.06.2026).

7. Grootendorst M. KeyBERT: Minimal Keyword Extraction with BERT [Electronic resource] // Zenodo. — 2020. — URL: https://doi.org/10.5281/zenodo.4461265 (accessed: 19.06.2026).

8. Sinha N. QA-prompting: Improving Summarization with Language Model Question Answering [Electronic resource] // arXiv preprint arXiv:2505.14347. — 2025. — URL: https://doi.org/10.48550/arXiv.2505.14347 (accessed: 19.06.2026).

9. Pang J., Ye F., Wong D.F., He X., Chen W., Wang L. Anchor-based Large Language Models [Electronic resource] // arXiv preprint arXiv:2402.07616. — 2024. — URL: https://doi.org/10.48550/arXiv.2402.07616 (accessed: 19.06.2026).

10. Kaplan J., McCandlish S., Henighan T., Brown T.B., Chess B., Child R. et al. Scaling Laws for Neural Language Models [Electronic resource] // arXiv preprint arXiv:2001.08361. — 2020. — URL: https://doi.org/10.48550/arXiv.2001.08361 (accessed: 19.06.2026).

11. Solaiman I., Brundage M., Clark J., Askell A., Herbert-Voss A., Wu J. et al. Release Strategies and the Social Impacts of Language Models [Electronic resource] // arXiv preprint arXiv:1908.09203. — 2019. — URL: https://doi.org/10.48550/arXiv.1908.09203 (accessed: 19.06.2026).

12. Lewis M., Liu Y., Goyal N., Ghazvininejad M., Mohamed A., Levy O. et al. BART: Denoising Sequence-to-Sequence Pre-training for Natural Generation, Translation, and Comprehension [Electronic resource] // arXiv preprint arXiv:1910.13461. — 2019. — URL: https://doi.org/10.48550/arXiv.1910.13461 (accessed: 19.06.2026).

13. Sanh V., Debut L., Chaumond J., Wolf T. DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter [Electronic resource] // arXiv preprint arXiv:1910.01108. — 2019. — URL: https://doi.org/10.48550/arXiv.1910.01108 (accessed: 19.06.2026).

14. DeepSeek-AI, Liu A., Feng B., Xue B., Wang B., Wu B. et al. DeepSeek-V3 Technical Report [Electronic resource] // arXiv preprint arXiv:2412.19437. — 2024. — URL: https://doi.org/10.48550/arXiv.2412.19437 (accessed: 19.06.2026).

УДК 004.67:004.93:007.52

Спатаев Айбар

Магистрант 2 курса
кафедра информационных технологий
Северо-Западный Политехнический Университет
(г. Сиань, КНР)
Научный руководитель: Сатымбеков М.Н.
PhD, цифровой офицер департамента
Университет Шакарима
(г. Семей, Казахстан)

ПРОГРАММНАЯ СИМУЛЯЦИЯ ДЕГРАДАЦИИ LIDAR В УСЛОВИЯХ ЗАДЫМЛЕНИЯ ДЛЯ АВТОНОМНЫХ АГЕНТОВ В КИБЕР-ФИЗИЧЕСКИХ СИСТЕМАХ

Аннотация: Надёжная автономная работа воплощённых агентов в условиях катастроф представляет критическую задачу для кибер-физических систем совместной работы человека и машины. В данной работе исследуется деградация точности LiDAR-сенсоров при высокой оптической плотности среды — в частности, дыма от пожара. Существующие симуляторы не воспроизводят физику аэрозольного рассеивания, препятствуя реалистичной проверке алгоритмов пространственной фильтрации. Предложена гибридная программная модель, сочетающая процедурную генерацию оптической плотности среды методом шума Перлина с вероятностным алгоритмом трассировки лучей на основе закона Бугера–Ламберта–Бера. Модель генерирует физически правдоподобные «фантомные препятствия» и бимодальные распределения интенсивности в реальном времени.

Ключевые слова: симуляция LiDAR, воплощённые агенты, кибер-физические системы, рассеивание дыма, шум Перлина

Введение. Концепция всепроникающих вычислений предполагает бесшовное встраивание интеллектуальных систем в физическую среду для поддержки деятельности человека в режиме реального времени — в том числе в условиях, опасных для прямого присутствия человека [1]. Сценарии ликвидации катастроф, такие как пожар в здании, представляют собой парадигматический случай кибер-физического взаимодействия человека и машины: автономные воплощённые агенты должны надёжно функционировать в деградированных средах, расширяя ситуационную осведомлённость спасательных команд [8, 18].

Фундаментальным требованием для таких систем является робастная одновременная локализация и картографирование (SLAM), традиционно реализуемая с помощью LiDAR-сенсоров [6, 7, 21, 22]. Однако эффективность LiDAR-систем критически снижается в средах с высокой аэрозольной оптической плотностью. При взаимодействии лазерного импульса с частицами дыма происходят два основных явления: ослабление сигнала [11] и ложные отражения вследствие эффектов рассеивания Ми [9, 13, 14], формирующие непроходимые «фантомные препятствия» в облаке точек.

Эта сбойность непосредственно распространяется на контур взаимодействия человека и машины: когда алгоритмы навигации Nav2 в среде ROS 2 интерпретируют дым как сплошную стену, воплощённый агент останавливается, лишая оператора надёжных ситуационных данных в наиболее критический момент [1, 24]. Проведение натурных испытаний в реальной пожарной среде экономически нецелесообразно и сопряжено с высоким риском, что обусловливает потребность в программных симуляторах, совместимых с ресурсно-ограниченным граничным оборудованием всепроникающих вычислений.

Обзор существующих решений. Для выявления существующих подходов к программному моделированию работы LiDAR в оптически плотных средах был проведён систематический обзор литературы в базах данных Scopus, Web of Science и IEEE Xplore за период с 2018 по 2026 год. Первоначальный поиск выявил 145 потенциально релевантных записей; после скрининга 32 статьи были отобраны для полнотекстового анализа, из которых 12 ключевых исследований вошли в итоговый синтез.

Синтез выявил два основных подхода. Базовое трассирование лучей с визуальными шейдерами (Gazebo, Webots, CARLA, AirSim) обнаруживает пересечения только с полигональными сетками [2, 19, 20], полностью исключая генерацию ложного обратного рассеивания. Стохастическое и объёмное моделирование (Монте-Карло, реймаршинг) обеспечивает физически точные модели, однако приводит к критическому падению производительности ниже 5 FPS [12], что несовместимо с требованиями реального времени.

Цель исследования и вклад работы. Основная цель данной работы — преодолеть разрыв между вычислительно затратными физическими моделями аэрозолей и чрезмерно упрощёнными визуальными шейдерами. Конкретный вклад включает: (1) вычислительно лёгкий программный инструментарий для симуляции деградации LiDAR с использованием шума Перлина [5]; (2) вероятностный алгоритм трассировки лучей на основе закона Бугера–Ламберта–Бера [15]; (3) количественную оценку вычислительной эффективности и верификацию распределений интенсивности.

Методология гибридной программной модели. Архитектура построена с чётким разделением двух взаимосвязанных компонентов: процедурной генерации аэрозольной оптической плотности и вероятностной модели трассировки лучей LiDAR-сенсора.

Процедурная генерация оптической плотности среды. Локальная оптическая плотность D в произвольной пространственной точке вычисляется посредством суперпозиции октав шума Перлина:

$$D(x,y,z) = max\{0, \Sigma A_i \cdot noise(f_i \cdot x, f_i \cdot (y - v \cdot t), f_i \cdot z) + C\} \quad (1)$$

где A_i и f_i — амплитуда и частота i-й октавы, v — скорость конвективного потока, t — время симуляции, C — константа смещения. Результирующее поле оптической плотности и его контуры представлены на Рисунке 1.

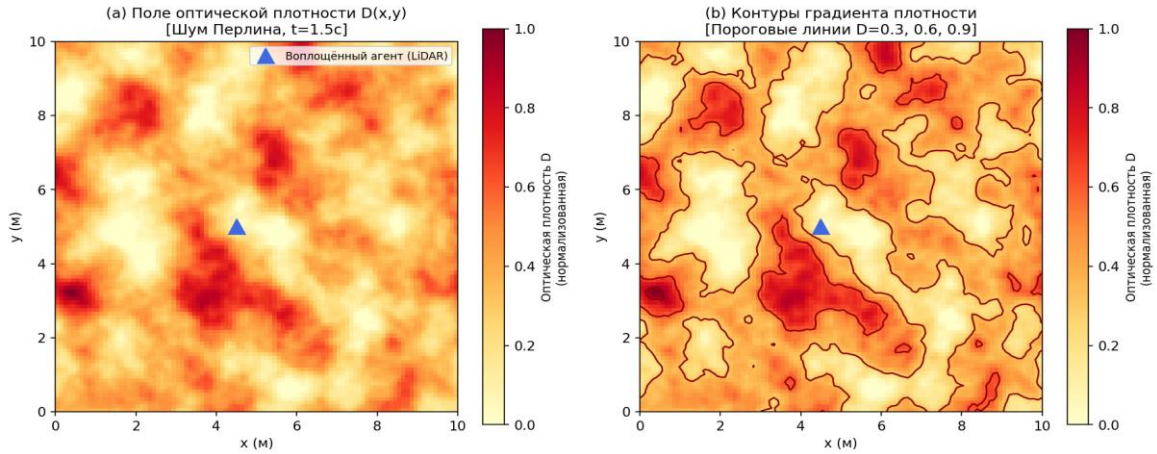


Рисунок 1 – Процедурно сгенерированное поле аэрозольной оптической плотности: (a) поле D(x,y) при t=1.5c, (b) контуры градиента плотности с пороговыми линиями D=0.3, 0.6, 0.9

Вероятностная модель деградации данных LiDAR. Полная логика выполнения алгоритма за кадр представлена на Рисунке 2.

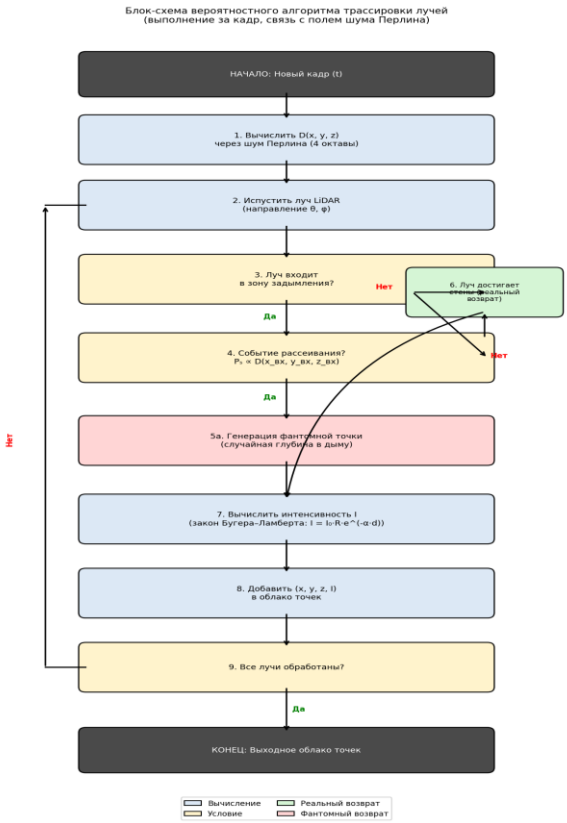


Рисунок 2 – Блок-схема вероятностного алгоритма трассировки лучей LiDAR (выполнение за кадр)

При испускании лазерного импульса вероятность P_s рассеивания луча прямо пропорциональна локальной аэрозольной плотности:

$$Pscatter \propto D(x\_entry, y\_entry, z\_entry) \quad (2)$$

Интенсивность отражённого сигнала I вычисляется на основе закона Бугера–Ламберта–Бера:

$I = I_0 \cdot R \cdot e^{(-\alpha \cdot D \cdot d)} / (1 + \beta \cdot d)$ (3)

где I_0 — базовая мощность излучателя, R — альбедо, α — коэффициент экстинкции, d — расстояние, β — коэффициент расходимости луча.

Экспериментальные результаты. Модель оценивалась в симулированной среде 10×10×10 м на стандартном CPU (Intel Core i5, 8 ГБ ОЗУ) без GPU-ускорения, с использованием NumPy [3]. LiDAR-сенсор конфигурировался с 360 лучами; параметры: I_0=1.0, R=0.9 (стены), R=0.15 (сажа), α=0.8, β=0.05. Шум Перлина генерировался с 4 октавами, персистентностью 0.5, лакунарностью 2.0.

Сравнение с существующими методами. В Таблице 1 представлено сопоставление предложенной модели с существующими подходами.

Таблица 1 – Сравнение методов симуляции LiDAR в дыму

Метод	FPS	Обратное рассеивание	Граничный CPU	Физическая адекватность
Шейдер Gazebo	60+	Нет	Да	Низкая
Монте-Карло / Фотон	<5	Да	Нет	Высокая
Полный реймаршинг	<5	Да	Нет	Высокая
Предложенная гибридная модель	30+	Да	Да	Средне-высокая

Вычислительная производительность. Предложенная модель поддерживала стабильную рабочую частоту 30+ FPS при всех значениях оптической плотности (D = 0 до 1.0), что демонстрирует Рисунок 3. В отличие от полного реймаршинга, деградирующего ниже 5 FPS, производительность гибридной модели оставалась стабильной даже при максимальной плотности дыма.

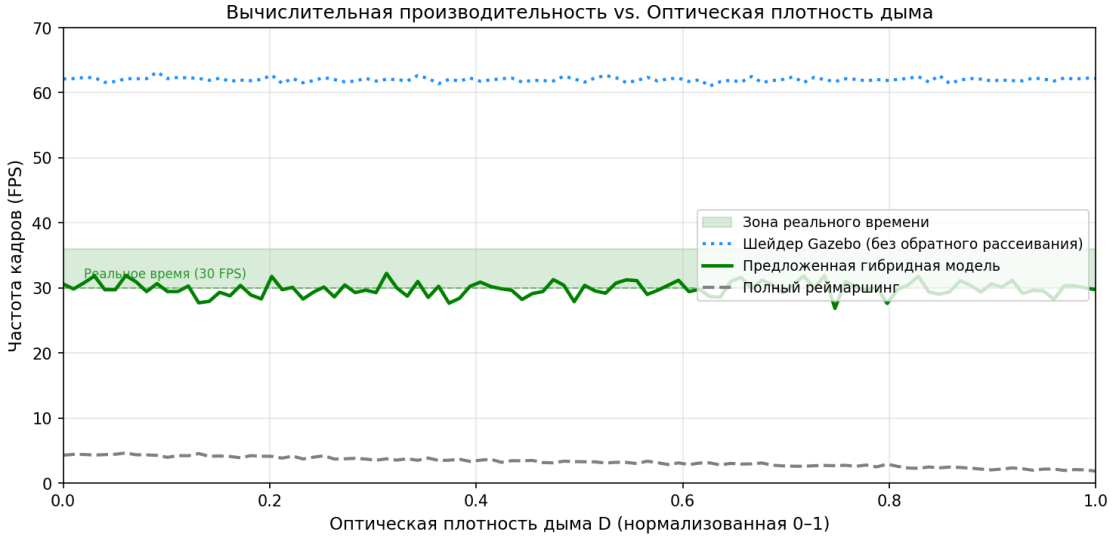


Рисунок 3 – Вычислительная производительность (FPS) vs. оптическая плотность дыма D

Анализ распределения интенсивности. Количественный анализ подтверждает физическую адекватность модели. Бимодальное распределение (Рисунок 4) чётко

разделяет две популяции: возвраты от стен (среднее ≈ 0.85, σ ≈ 0.06) и аэрозольные фантомные точки (среднее ≈ 0.15, σ ≈ 0.09), что коррелирует с предсказанием закона Бугера–Ламберта–Бера при D = 0.7.

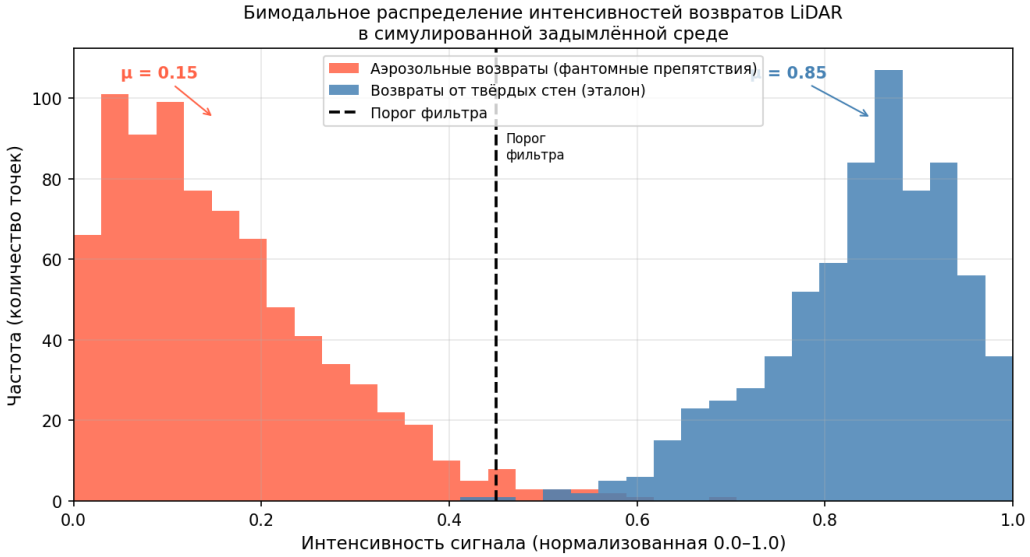


Рисунок 4 – Бимодальное распределение интенсивностей возвратов LiDAR в симулированной задымлённой среде

Верификация закона Бугера–Ламберта–Бера. На Рисунке 5 представлено сопоставление теоретических кривых с симулированными измерениями облака точек при D=0.7. Результаты подтверждают корректность применения физической модели.

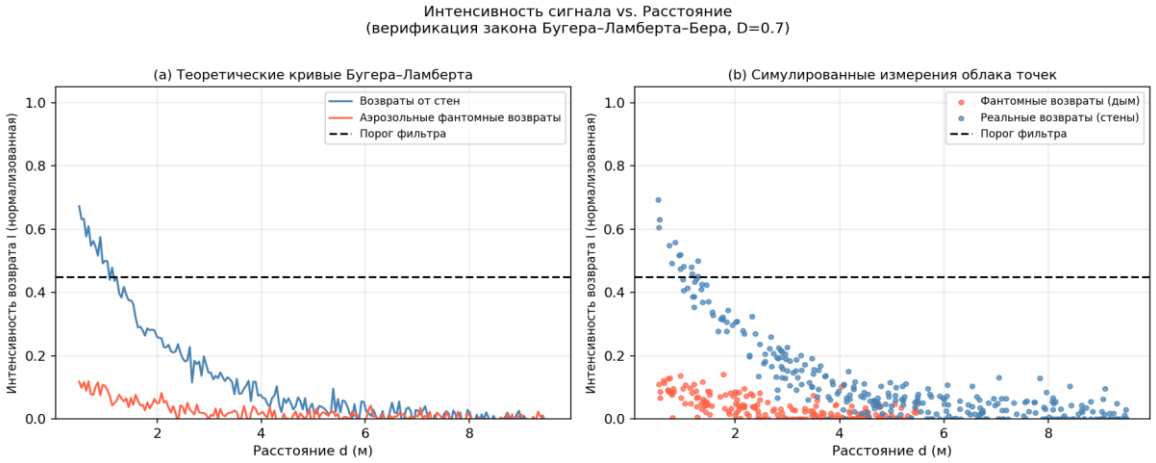


Рисунок 5 – Интенсивность сигнала vs. расстояние: (a) теоретические кривые Бугера–Ламберта, (b) симулированные измерения облака точек (D=0.7)

Визуализация облака точек. Рисунок 6 демонстрирует различие между сырым облаком точек (с фантомными препятствиями, блокирующими путь Nav2) и отфильтрованным облаком после применения фильтра SOR на основе интенсивности.

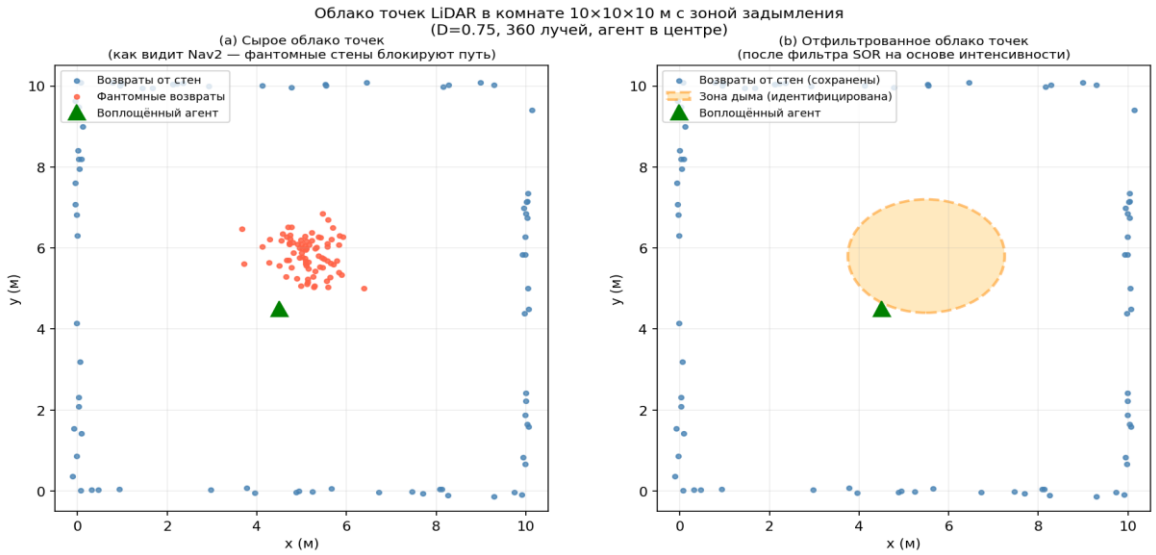


Рисунок 6 – Облако точек LiDAR в комнате 10×10×10 м: (a) сырые данные с фантомными препятствиями, (b) после фильтрации с идентификацией зоны дыма

Доля фантомных препятствий. Рисунок 7 подтверждает нелинейный рост доли фантомных точек с увеличением плотности дыма. Симулированные измерения хорошо согласуются с теоретической кривой Бугера–Ламберта–Бера при всех режимах задымления.

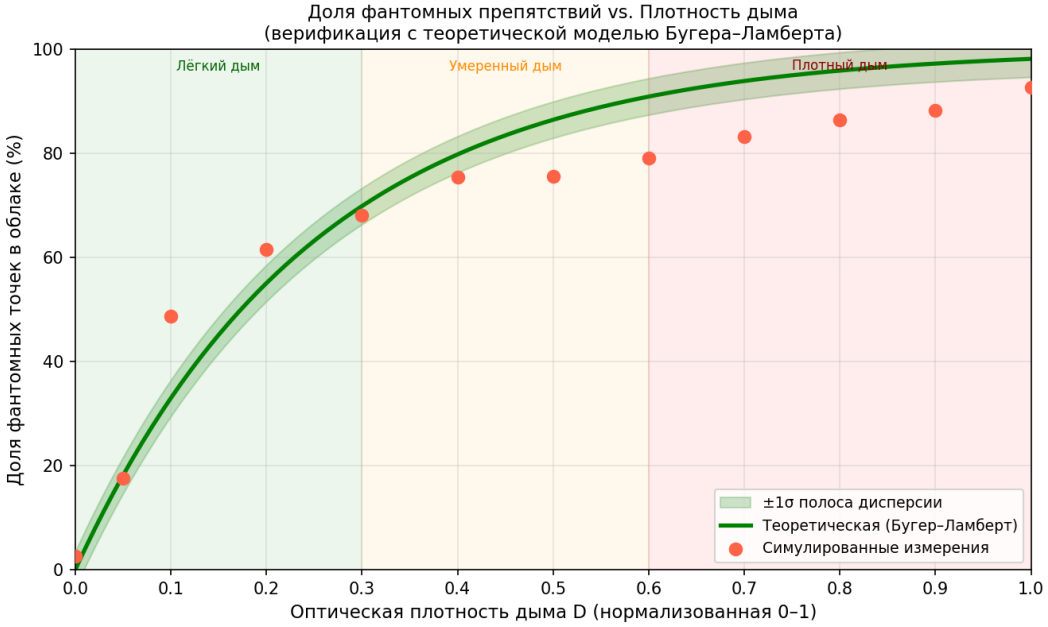


Рисунок 7 – Доля фантомных препятствий vs. плотность дыма (верификация с теоретической моделью Бугера–Ламберта)

Обсуждение. Разработанная гибридная модель устраняет ограничения, выявленные в ходе систематического обзора. В отличие от стандартных визуальных шейдеров, модель физически искажает конвейер пространственных данных, вынуждая алгоритмы навигации сталкиваться с теми же аномалиями, что и в реальных пожарных сценариях. Замена полного реймаршинга вероятностным граничным пороговым приближением обеспечивает стабильную частоту кадров (>30 FPS) на стандартном CPU-оборудовании.

Предлагаемый симулятор позволяет проводить реалистичное SITL-тестирование пространственных фильтров — таких как статистическое удаление выбросов (SOR) и пороговая обработка на основе интенсивности, — помогая закрыть валидационный разрыв в контуре взаимодействия человека и машины [16, 23, 25]. Основные ограничения текущей модели: упрощение расходимости луча и многократных эхо-отражений; аппроксимация шумом Перлина не полностью воспроизводит временную динамику движения частиц дыма. Будущие работы расширят модель на 3D-конфигурации LiDAR и интеграцию ROS 2.

Заключение. В данной работе успешно разработана и оценена программная модель для симуляции воздействия плотного дыма на точность данных LiDAR в кибер-физических средах ликвидации катастроф. Сочетание процедурной генерации шума Перлина с вероятностным применением закона Бугера–Ламберта–Бера позволяет симулятору генерировать физически правдоподобные «фантомные препятствия» и вариации интенсивности в реальном времени при производительности свыше 30 FPS на стандартном CPU. Предложенная модель обеспечивает безопасный, ресурсоэффективный и физически адекватный виртуальный испытательный стенд для разработки и верификации алгоритмов пространственной фильтрации для воплощённых агентов в сценариях реагирования на катастрофы.

Список литературы:

1. Bijelic, M., Gruber, T., et al.: Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In: Proceedings of the IEEE/CVF CVPR (2020)
2. Bohren, C.F., Huffman, D.R.: Absorption and Scattering of Light by Small Particles. Wiley, Hoboken (2008)
3. Cadena, C., Carlone, L., et al.: Past, present, and future of simultaneous localization and mapping. IEEE Trans. Robot. 32(6), 1309–1332 (2016)
4. Delmerico, J., Mintchev, S., et al.: The current state and future outlook of rescue robotics. J. Field Robot. 36(7), 1171–1191 (2019)
5. Dosovitskiy, A., Ros, G., et al.: CARLA: An open urban driving simulator. In: CoRL, pp. 1–16. PMLR (2017)
6. Goodin, C., Doude, M., et al.: Predicting the influence of rain on LIDAR in ADAS. Electronics 8(1), 89 (2019)
7. Gubbi, J., Buyya, R., et al.: Internet of Things (IoT): A vision, architectural elements, and future directions. Future Gener. Comput. Syst. 29(7), 1645–1660 (2013)
8. Hahner, M., Tremblay, M., et al.: Fog simulation on real lidar point clouds for 3D object detection in adverse weather. In: IEEE/CVF ICCV, pp. 3726–3737 (2021)
9. Harris, C.R., Millman, K.J., et al.: Array programming with NumPy. Nature 585(7825), 357–362 (2020)
10. Hasirlioglu, S., Kamann, A., Dorndorf, U.: Test methodology for rain and fog effects on automotive lidar. In: IEEE IV Symposium, pp. 843–848. IEEE (2017)
11. Hui, Z., et al.: LIDSOR: A filter for removing rain and snow noise points from LiDAR point clouds. ISPRS J. Photogramm. (2023)
12. Kim, J.H., Latchman, H.A., et al.: Sensor fusion based seek-and-find fire algorithm for intelligent firefighting robot. IEEE Trans. (2018)

13. Koenig, N., Howard, A.: Design and use paradigms for Gazebo, an open-source multi-robot simulator. In: IEEE/RSJ IROS, vol. 3, pp. 2149–2154. IEEE (2004)
14. Kummerle, R., Grisetti, G., et al.: g2o: A general framework for graph optimization. In: IEEE ICRA, pp. 3607–3613. IEEE (2011)
15. Macenski, S., Foote, T., et al.: Robot Operating System 2: Design, architecture, and uses in the wild. Sci. Robot. 7(66), eabm6074 (2022)
16. Perlin, K.: An image synthesizer. ACM SIGGRAPH Comput. Graph. 19(3), 287–296 (1985)
17. Pomerleau, F., Colas, F., et al.: Comparing ICP variants on real-world data sets. Auton. Robots 34(3), 133–148 (2013)
18. Rasshofer, R.H., Spies, M., Spies, H.: Influences of weather phenomena on automotive laser radar systems. Adv. Radio Sci. 9, 49–60 (2011)
19. Rusu, R.B., Cousins, S.: 3D is here: Point Cloud Library (PCL). In: IEEE ICRA, pp. 1–4. IEEE (2011)
20. Shah, S., Dey, D., et al.: AirSim: High-fidelity visual and physical simulation for autonomous vehicles. In: Field and Service Robotics. Springer (2017)
21. Shan, T., Englot, B.: LeGO-LOAM: Lightweight and ground-optimized lidar odometry and mapping. In: IEEE/RSJ IROS, pp. 4758–4765. IEEE (2018)
22. Shi, W., Cao, J., et al.: Edge computing: Vision and challenges. IEEE Internet Things J. 3(5), 637–646 (2016)
23. Swinehart, D.F.: The Beer-Lambert law. J. Chem. Educ. 39(7), 333 (1962)
24. Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics. MIT Press, Cambridge (2005)
25. Weiser, M.: The computer for the 21st century. Sci. Am. 265(3), 94–104 (1991)
26. Zhang, J., Singh, S.: LOAM: Lidar odometry and mapping in real-time. In: RSS, vol. 2 (2014)
27. Zheng, et al.: Physically-based LiDAR smoke simulation for robust 3D object detection. AAAI Publications (2026)
28. Zhou, Y., et al.: Fire SLAM: A visual SLAM algorithm for firefighting robots. ResearchGate (2024)

МЕДИЦИНАЛЫҚ ҒЫЛЫМДАР - МЕДИЦИНСКИЕ НАУКИ - MEDICAL SCIENCES

УДК 338.262.4

Касымова Орынтай Давлетпаевна

Докторант DBA,
Алматы Менеджмент Университет
(г. Алматы, Казахстан)

ПРЕИМУЩЕСТВА И ЭФФЕКТИВНОСТЬ ТЕЛЕМЕДИЦИНЫ ПЕРЕД ТРАДИЦИОННЫМИ МЕТОДАМИ

Аннотация: В исследовании подчеркиваются политические последствия внедрения телемедицины, включая необходимость нормативно-правовой базы, политики возмещения расходов и этических соображений. Также был отмечен потенциал телемедицины в решении проблем неравенства в здравоохранении и содействии равноправному доступу к медицинским услугам. В заключение исследования рекомендуется проведение дальнейших исследований и политических инициатив для повышения эффективности использования телемедицины в сельских и малообеспеченных районах. Данная исследовательская работа вносит вклад в существующий объем знаний, предоставляя всесторонний анализ потенциала телемедицины в улучшении оказания медицинской помощи в малообеспеченных сообществах.

Ключевые слова: финансовое управление в здравоохранении, управление медицинскими учреждениями, снижение затрат медицинских учреждений, операционное управление медицинскими учреждениями

Телемедицина в своей сущности не является новым явлением само по себе. Она практикуется с использованием электронных средств связи уже десятилетиями, а с использованием традиционных форм коммуникации — гораздо дольше. Несмотря на этот исторический контекст, телемедицина, похоже, оказывает поляризующее воздействие на профессию врача. Люди редко занимают нейтральную позицию по этому вопросу; они либо являются ярыми сторонниками, либо яростными противниками. Сторонники считают, что телемедицина представляет собой будущее. Она приведет к повышению стандартов медицинской помощи, а также к снижению затрат. Противники считают, что она представляет угрозу традиционным отношениям врача и пациента и является по своей сути небезопасным способом медицинской практики. Потенциальные юридические и этические проблемы, связанные с телемедициной, часто используются как «завеса», чтобы поддержать мнение о том, что возможные осложнения телемедицины означают, что ее нельзя использовать в качестве основы для клинической службы [1].

Телемедицина стала универсальным инструментом в оказании медицинской помощи, предлагая множество применений, способных произвести революцию в уходе за пациентами. В этом разделе рассматриваются три ключевых применения телемедицины: дистанционные консультации с врачами, выписка рецептов и дистанционный мониторинг состояния пациентов.

Было выявлено множество успешных программ телемедицины, демонстрирующих улучшение доступа к медицинским услугам в отдаленных и малообеспеченных районах. Преимущества телемедицины проявлялись в расширении доступа к медицинской помощи, снижении затрат и повышении удовлетворенности пациентов. Однако были также выявлены проблемы, такие как ограниченная технологическая инфраструктура, нормативные барьеры и проблемы конфиденциальности.

Телемедицина предлагает множество применений, улучшающих оказание медицинской помощи. Дистанционные консультации с врачами преодолевают географические барьеры, обеспечивают своевременный доступ к специализированной помощи и оптимизируют использование ресурсов здравоохранения. Выписка рецептов с помощью телемедицины обеспечивает удобство, экономию средств и улучшение соблюдения режима приема лекарств. Дистанционный мониторинг состояния пациентов позволяет непрерывно отслеживать их здоровье, улучшает управление заболеваниями и поддерживает профилактическую медицинскую помощь. Используя потенциал телемедицинских приложений, системы здравоохранения могут расширить доступ к медицинской помощи, улучшить результаты лечения пациентов и трансформировать систему оказания медицинских услуг [2].

Дистанционные консультации позволяют пациентам связываться с медицинскими работниками и специалистами с помощью телекоммуникационных технологий, устраняя необходимость в личных визитах. С помощью видеоконференций или телефонных разговоров пациенты могут обсуждать свои медицинские проблемы, получать медицинские консультации и обращаться за экспертным мнением к медицинским специалистам [3]. Это применение телемедицины доказало свою эффективность в нескольких медицинских специальностях, включая первичную медицинскую помощь, дерматологию, психиатрию и радиологию.

Дистанционные консультации предоставляют ряд преимуществ. Во-первых, они преодолевают географические барьеры, позволяя пациентам в сельских и малообеспеченных районах получать специализированную помощь, которая может быть недоступна на местном уровне. Во-вторых, они снижают транспортные издержки и связанные с ними расходы для пациентов, особенно для тех, кто проживает в отдаленных регионах. Кроме того, дистанционные консультации экономят время, устраняя необходимость ожидания в медицинских учреждениях, обеспечивая удобный и своевременный доступ к медицинской экспертизе [2]. Эти преимущества улучшают доступ к медицинским услугам, повышают удовлетворенность пациентов и оптимизируют использование ресурсов здравоохранения.

Телемедицина упростила доставку рецептов пациентам дистанционно, обеспечивая удобство и эффективность. С помощью телекоммуникационных каналов медицинские работники могут удаленно просматривать медицинские карты пациентов, оценивать симптомы и выписывать лекарства. Пациенты могут получать свои рецепты в электронном виде, которые могут быть напрямую пересланы в аптеки или доставлены почтой.

Доставка рецептов посредством телемедицины имеет ряд преимуществ. Она устраняет необходимость в личных визитах только для пополнения запасов лекарств, экономя время пациентов и сокращая транспортные расходы. Она также способствует соблюдению режима приема лекарств, обеспечивая своевременный доступ к лекарствам,

минимизируя перерывы в лечении и снижая вероятность несоблюдения режима приема лекарств [4]. Кроме того, доставка рецептов посредством телемедицины может быть особенно полезна для людей с ограниченной подвижностью или для тех, кто проживает в отдаленных районах с ограниченным доступом к местным аптекам.

Дистанционный мониторинг состояния пациентов предполагает использование технологий для отслеживания состояния здоровья пациентов, жизненно важных показателей, симптомов и прогресса лечения на расстоянии. Это применение телемедицины позволяет медицинским работникам отслеживать и оценивать состояние здоровья пациентов удаленно, сводя к минимуму необходимость частых личных визитов.

Дистанционный мониторинг состояния пациентов использует различные устройства, такие как носимые датчики, мобильные приложения для здравоохранения и оборудование для телемониторинга. Эти устройства позволяют собирать и передавать данные в режиме реального времени, что позволяет медицинским работникам отслеживать состояние пациентов и вмешиваться при необходимости. Это приложение особенно полезно для лечения хронических заболеваний, таких как диабет, гипертония и болезни сердца.

Дистанционный мониторинг состояния пациентов предлагает ряд преимуществ. Он улучшает управление заболеваниями, предоставляя медицинским работникам непрерывные объективные данные о состоянии здоровья пациентов, что позволяет своевременно вмешиваться и корректировать планы лечения. Пациенты получают выгоду от сокращения числа госпитализаций, улучшения самоконтроля и повышения вовлеченности в собственное здравоохранение. Кроме того, дистанционный мониторинг состояния пациентов может способствовать раннему выявлению осложнений или изменений в состоянии здоровья, что приводит к проактивному и профилактическому уходу [5].

Однако важно понимать ограничения дистанционного мониторинга состояния пациентов. Он может быть непригоден для всех медицинских состояний, поскольку некоторые состояния могут потребовать очных осмотров или вмешательств. Кроме того, успешная реализация дистанционного мониторинга зависит от готовности пациентов активно участвовать, предоставлять точные данные и соблюдать протоколы мониторинга.

Телемедицина стала революционным решением для преодоления разрыва в доступе к медицинской помощи и улучшения результатов лечения в сельских и малообеспеченных районах. Используя технологии и инновационные коммуникационные платформы, телемедицина предлагает ряд преимуществ, позволяющих решать уникальные проблемы, с которыми сталкиваются эти группы населения.

Телемедицина является жизненно важным инструментом для людей, проживающих в отдаленных или географически изолированных районах. Она преодолевает географические барьеры, связывая пациентов с медицинскими работниками независимо от их физического местоположения. Пациентам больше не нужно преодолевать большие расстояния или долго ждать медицинской помощи. Благодаря телемедицине они могут получить доступ к широкому спектру медицинских услуг, включая консультации, диагностические обследования, последующие визиты и постоянное управление лечением. Этот расширенный доступ к медицинским услугам

позволяет своевременно оказывать помощь, сокращает неравенство в здравоохранении и способствует улучшению результатов лечения в сельских и малообеспеченных районах [6].

Расширение доступа к специализированной медицинской помощи: Специализированная медицинская помощь часто является дефицитной или отсутствует в сельских и малообеспеченных районах. Телемедицина облегчает доступ к квалифицированным медицинским работникам, позволяя проводить дистанционные консультации и виртуальные визиты. Пациенты могут связаться со специалистами в различных областях, таких как кардиология, дерматология, психиатрия и радиология, без необходимости преодолевать большие расстояния. Этот доступ к специализированной экспертизе повышает качество медицинской помощи, улучшает точность диагностики и позволяет своевременно давать рекомендации по лечению. Благодаря телемедицине сельские и малообеспеченные сообщества получают доступ к более широкому спектру медицинских знаний и опыта, что в конечном итоге улучшает результаты лечения и удовлетворенность пациентов [7].

Телемедицина способствует своевременному вмешательству и профилактической помощи, что крайне важно в сельских и малообеспеченных районах, где доступ к здравоохранению может быть ограничен. Благодаря дистанционному мониторингу медицинские работники могут отслеживать жизненно важные показатели пациентов, симптомы и прогрессирование заболевания в режиме реального времени. Это позволяет выявлять осложнения на ранней стадии, проводить упреждающие вмешательства и профилактические меры. Для людей с хроническими заболеваниями телемедицина предлагает непрерывное управление заболеванием, позволяя пациентам активно участвовать в собственном лечении и принимать обоснованные решения об изменении образа жизни и планах лечения. Своевременное вмешательство и профилактическая помощь, обеспечиваемые телемедициной, способствуют улучшению показателей здоровья, снижению числа госпитализаций и улучшению общего состояния здоровья населения в сельских и малообеспеченных районах [4].

Телемедицина способствует расширению прав и возможностей пациентов и их вовлечению в принятие решений в сфере здравоохранения. Предоставляя удобный доступ к медицинским услугам, пациенты получают больший контроль над своим здоровьем и могут активно участвовать в управлении своим состоянием. Платформы телемедицины часто предлагают образовательные ресурсы для пациентов, инструменты удаленного мониторинга и защищенные системы обмена сообщениями, которые способствуют общению между пациентом и врачом и повышают вовлеченность пациентов. Пациенты, обладающие расширенными правами и возможностями, с большей вероятностью будут соблюдать планы лечения, вести более здоровый образ жизни и играть активную роль в управлении своим здоровьем, что в конечном итоге приводит к улучшению показателей здоровья в сельских и малообеспеченных районах [8].

В заключение телемедицина предлагает значительные преимущества и оказывает существенное влияние на улучшение доступа к медицинской помощи и результатов лечения в сельских и малообеспеченных районах. Расширяя доступ к медицинским услугам, обеспечивая специализированную помощь, снижая затраты, позволяя своевременно проводить вмешательства и профилактическую помощь, а также способствуя расширению прав и возможностей пациентов и их вовлечению,

телемедицина решает уникальные проблемы, с которыми сталкиваются эти группы населения. Используя возможности технологий, телемедицина имеет потенциал для преобразования системы здравоохранения, сокращения неравенства и улучшения общего состояния здоровья и благополучия людей в сельских и малообеспеченных общинах.

Список литературы:

1. Медведева Е. И., Александрова О. А., Крошилин С. В. Телемедицина в современных условиях: отношение социума и вектор развития // Экономические и социальные перемены: факты, тенденции, прогноз. 2022. Т. 15, № 3. С. 200–222.
2. Петреева А.С., Казарян И.Р. Телемедицина — новые возможности в здравоохранении. «Аспирант». Приложение к журналу «Вестник
3. Городнова Н. В., Клевцов В. В., Овчинников Е. Н. Перспективы развития телемедицины в условиях цифровизации экономики России // Вопросы инновационной экономики. – 2019. – Т. 9, № 3. – С. 1049–1066. – DOI: 10.18334/vinec.9.3.41173.
4. Федяев Д. В., Федяева В. К., Омельяновский В. В. Экономическое обоснование применения телемедицинских технологий для диспансеризации населения в отдаленных районах // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2014. – Т. 7, № 3. – С. 30–35.
5. Губина О. В. Формирование телемедицинской системы как инновационного фактора развития арктических территорий России // Региональные проблемы преобразования экономики. – 2020. – № 5 (115). – С. 39–47. – DOI: 10.26726/1812-7096-2020-5-39-47.
6. Владзимирский А.В., Морозов С.П., Сименюра С.С. (2020). Телемедицина и COV-19: оценка качества телемедицинских консультаций, инициированных пациентами с симптомами ОРВИ // Врач и информационные технологии. № 2. С. 52—63. БО1: 10.37690/1811-0193-2020-2-52-63
7. Морозов С.П., Владзимирский А.В., Сименюра С.С. (2020). Качество первичных телемедицинских консультаций «пациент-врач» (по результатам тестирования телемедицинских сервисов) // Врач и информационные технологии. № 1. С. 51—62.
8. Шадеркин И. А. Барьеры телемедицины и пути их преодоления // Экспериментальная и клиническая урология. – 2022. – № 1. – С. 160–164.

ПЕДАГОГИКА ЖӘНЕ ПСИХОЛОГИЯ ҒЫЛЫМДАР - ПЕДАГОГИЧЕСКИЕ И ПСИХОЛОГИЧЕСКИЕ НАУКИ - PEDAGOGICAL AND PSYCHOLOGICAL SCIENCES

UDC 347.1

Aldabergen Gulsim Erkinkyzy,

Suleeva Madina

Master's students
South Kazakhstan Pedagogical University
named after Ozbekali Zhanibekov
(Shymkent, Kazakhstan)

HARNESSING WEB QUEST TECHNOLOGY TO ENHANCE FOREIGN LANGUAGE COMMUNICATIVE COMPETENCE: A REVIEW OF STRATEGIES AND IMPACTS FOR LANGUAGE UNIVERSITY STUDENTS

Abstract: Web Quest technology has emerged as an innovative pedagogical tool for enhancing foreign language communicative competence among language university students. This approach integrates digital resources with structured, inquiry-based tasks, enabling learners to engage with authentic and meaningful content while practicing language skills in real-world contexts. As the demand for effective, technology-driven teaching strategies grows, understanding the potential of Web Quests in fostering linguistic, pragmatic, and collaborative competencies has become increasingly relevant. Unlike traditional methods, which often lack engagement and contextual application, Web Quests provide an interactive platform that aligns with modern communicative language teaching principles.

The primary aim of this study is to evaluate the role of Web Quest technology in developing foreign language communicative competence. A systematic literature review was conducted, focusing on studies that explore the benefits, challenges, and practical applications of Web Quests in language education. Key findings reveal that Web Quests enhance student engagement and motivation through interactive and task-based learning. They improve linguistic accuracy, discourse cohesion, and pragmatic skills by immersing students in authentic communication scenarios. Additionally, the collaborative nature of Web Quests fosters teamwork, interpersonal communication, and critical thinking, all of which are integral to effective language use.

However, the review also identifies challenges associated with Web Quest implementation. These include the need for extensive teacher training, time-consuming task design, digital literacy disparities, and ensuring equitable access to technology. Addressing these challenges is essential for optimizing the potential of Web Quests as a didactic tool. The study underscores the importance of professional development programs for educators, structured task planning, and investment in digital infrastructure to ensure successful integration.

The significance of this research lies in its contribution to understanding how Web Quests can transform language learning by bridging the gap between theoretical knowledge and practical application. By providing students with immersive, technology-driven learning

experiences, Web Quests promote not only communicative competence but also essential 21st-century skills such as collaboration and problem-solving. The findings offer valuable insights for educators and curriculum designers seeking to innovate language education and prepare students for the communicative demands of a globalized world. This review concludes by recommending further research on long-term impacts, cross-cultural applications, and strategies for addressing implementation challenges to fully harness the potential of Web Quest technology in foreign language education.

Keywords: interactive learning, digital tools, task-based instruction, collaborative strategies, inquiry-based pedagogy, language proficiency, communicative skills

Introduction. The rapid advancement of digital technologies has revolutionized educational practices, offering innovative tools to enhance language learning. Among these, Web Quest technology has gained prominence as an interactive, task-based approach that integrates digital resources into the learning process. For language university students, developing foreign language communicative competence is a fundamental goal that requires a combination of linguistic, cognitive, and cultural skills. Web Quests, with their focus on inquiry-based learning and collaboration, provide a dynamic platform for achieving these competencies by immersing students in real-world, problem-solving tasks that simulate authentic communication scenarios.

The relevance of this topic lies in the growing demand for effective teaching methods that align with the needs of digitally native learners. Traditional approaches to communicative competence development often lack engagement and fail to fully utilize the potential of modern technology. Web Quests address these gaps by combining interactive activities, multimedia content, and authentic materials, fostering both motivation and deeper learning. Additionally, Web Quests align with the principles of learner autonomy and collaborative learning, empowering students to take active roles in their educational journey while enhancing their communicative skills in meaningful contexts.

The primary objective of this study is to explore the potential of Web Quest technology as a didactic means for developing foreign language communicative competence among language university students. The specific aims include:

- Evaluating the effectiveness of Web Quests in enhancing key components of communicative competence, such as linguistic accuracy, pragmatic skills, and cultural understanding.
- Analyzing existing research on the design and implementation of Web Quests in language education.
- Identifying challenges and best practices for integrating Web Quests into language university curricula.
- Proposing recommendations for educators and curriculum designers to optimize the use of Web Quests in fostering communicative competence.

By addressing these objectives, this research seeks to contribute to the development of innovative teaching practices that leverage digital tools to enhance language learning outcomes. The study underscores the importance of aligning educational strategies with technological advancements to prepare students for the communicative demands of a globalized world.

Theoretical background. Web Quest technology represents an innovative approach to language education, combining task-based learning principles with digital tools to create

engaging and authentic learning experiences. Web Quests, first introduced by Dodge (1995), are inquiry-based learning activities that guide students through a structured process of information gathering and problem-solving using web resources. Typically, Web Quests include components such as an introduction, task description, process steps, evaluation criteria, and a conclusion. In the context of language learning, Web Quests are designed to immerse students in real-world communicative scenarios that require them to use the target language for meaningful purposes. By incorporating multimedia elements and authentic materials, Web Quests provide rich contexts for developing both linguistic and pragmatic skills [1].

Communicative competence encompasses a range of skills necessary for effective language use, including grammatical competence (accuracy and structure), sociolinguistic competence (appropriate language use in context), discourse competence (coherence and cohesion), and strategic competence (managing communication breakdowns). According to Canale and Swain's (1980) model, these components are interrelated and essential for successful interaction in a foreign language. For language university students, communicative competence is a critical goal, as it prepares them for academic, professional, and intercultural communication [2].

Web Quests align with the principles of communicative language teaching (CLT) by emphasizing interaction, authentic communication, and learner-centered activities. Research indicates that Web Quests can effectively enhance various aspects of communicative competence. Linguistic accuracy is developed through tasks requiring precise language use, such as report writing or presentations. Pragmatic skills are fostered by exposing learners to culturally appropriate language use and scenarios. Collaborative learning is encouraged through group-based Web Quests, which promote teamwork and peer communication, fostering both language skills and social interaction. Inquiry-based tasks also promote critical thinking and analytical skills, which are essential for effective communication [3].

Several theoretical frameworks underpin the use of Web Quests in language education. Task-Based Language Teaching (TBLT) supports the use of Web Quests, as they involve goal-oriented activities requiring meaningful language use. Constructivist principles are evident, as Web Quests encourage learners to construct their own knowledge through exploration and interaction. Socio-cultural theory, based on Vygotsky's ideas, highlights the role of social interaction in learning, which is facilitated by collaborative problem-solving and peer learning in Web Quests [4].

Despite their benefits, Web Quests present challenges. Teachers may face difficulties in designing tasks that align with curriculum objectives and meet students' proficiency levels. Disparities in digital access and technical skills can hinder effective implementation, and the time-intensive nature of creating and managing Web Quests poses additional challenges. Nevertheless, the literature highlights that Web Quests hold significant potential for enhancing foreign language communicative competence, particularly for language university students. By integrating authentic tasks, multimedia resources, and collaborative activities, Web Quests address the limitations of traditional teaching methods while promoting engagement and learner autonomy. To maximize their impact, educators must focus on careful task design, training, and ensuring equitable access to digital tools [5]. This theoretical overview establishes Web Quest technology as a valuable didactic tool for language education, supported by strong theoretical foundations and practical evidence of its effectiveness in fostering communicative competence.

Table 1- Key aspects of using web quest technology for communicative competence

Aspect	Benefits	Challenges	Strategies
Engagement	Increases student motivation through interactive and inquiry-based tasks.	Designing tasks that balance engagement and language learning objectives.	Develop tasks that integrate authentic, relevant, and meaningful content.
Linguistic Accuracy	Encourages precise language use in structured tasks like report writing.	May not cater to all proficiency levels without proper scaffolding.	Provide pre-task activities and language support to help students prepare.
Pragmatic Skills	Enhances cultural understanding and appropriate language use in context.	Limited availability of culturally rich and diverse resources.	Curate high-quality, culturally appropriate materials and scenarios.
Collaboration	Promotes teamwork and peer communication through group tasks.	Managing group dynamics and ensuring equal participation.	Assign clear roles and responsibilities within teams to foster balanced collaboration.
Critical Thinking	Develops problem-solving and analytical skills through inquiry-based tasks.	Time-intensive preparation and task management for educators.	Use templates and pre-designed Web Quests to reduce preparation time while maintaining quality.
Technology Use	Provides access to multimedia resources and authentic web-based materials.	Disparities in digital access and technical skills among students.	Ensure equitable access to technology and provide training on digital literacy.

Methodology. This review employs a systematic approach to literature selection and thematic analysis to explore the role of Web Quest technology in developing foreign language communicative competence among language university students. The methodology ensures the inclusion of relevant, high-quality studies and a comprehensive synthesis of findings. The literature selection process involved several stages. First, major academic databases such as Google Scholar, ERIC, Scopus, and PubMed were searched to locate peer-reviewed journal articles, conference proceedings, and book chapters. Search terms including “Web Quest technology,” “foreign language communicative competence,” “task-based language teaching,” and “digital learning tools” were used in combination with Boolean operators to refine the results.

Studies were included based on specific inclusion criteria. They were required to focus on Web Quest technology or similar digital tools in language learning, address aspects of foreign language communicative competence such as linguistic accuracy, pragmatic skills, or collaboration, and present empirical evidence, theoretical frameworks, or practical recommendations. Articles were excluded if they lacked methodological rigor, focused on non-linguistic disciplines, or did not discuss Web Quest technology in the context of language education. Titles and abstracts of the identified studies were screened for relevance, and the full

texts of shortlisted articles were reviewed for final inclusion. The systematic search resulted in the selection of a diverse range of studies providing insights into the application of Web Quest technology in language education.

Once the relevant studies were selected, a thematic analysis was conducted to identify recurring patterns and key themes. This process involved systematically coding the findings of each study using qualitative analysis techniques. Codes such as “engagement,” “task-based learning,” “collaboration,” and “digital literacy” were applied to categorize the data. These codes were then grouped into broader themes, including the benefits of Web Quests for communicative competence, challenges in their implementation, and strategies for effective integration. Thematic analysis helped synthesize diverse perspectives, providing a comprehensive understanding of the role of Web Quests in language education.

This methodology ensured a robust framework for analyzing and interpreting the literature, highlighting both the potential and limitations of Web Quest technology in fostering foreign language communicative competence. By systematically identifying key insights and practical implications, this review aims to contribute to the development of effective teaching practices in language university settings.

Findings and Discussion. The findings from the literature review provide comprehensive insights into the potential of Web Quest technology as a didactic tool for developing foreign language communicative competence among language university students. Key themes emerged regarding the benefits of Web Quests, challenges in their implementation, and their implications for educational practice.

The review highlights that Web Quest technology significantly enhances student engagement by integrating interactive, inquiry-based tasks into the learning process. The structured nature of Web Quests, combined with their reliance on authentic web-based materials, fosters active participation and curiosity. Tasks such as collaborative problem-solving and simulated real-world scenarios provide students with opportunities to engage with the target language in meaningful ways, boosting both intrinsic motivation and sustained interest.

Web Quests effectively develop various aspects of communicative competence, including linguistic, pragmatic, and discourse skills. Linguistic accuracy is enhanced through tasks that require precise language use, such as written reports or oral presentations. Pragmatic skills are cultivated as students navigate culturally appropriate expressions and contextual language use in authentic scenarios. Additionally, discourse competence is improved through activities that promote coherence and cohesion, such as group discussions and collaborative writing tasks.

Web Quests encourage collaborative learning by involving students in group tasks that require teamwork and peer interaction. This collaborative approach not only improves language proficiency but also develops interpersonal communication skills, which are essential for professional and academic settings. Moreover, inquiry-based tasks challenge students to analyze, evaluate, and synthesize information, fostering critical thinking and problem-solving abilities.

Despite their benefits, Web Quests present notable challenges. Designing effective Web Quests that align with curriculum objectives and language proficiency levels can be time-intensive for educators. Technical barriers, such as disparities in digital access and varying levels of student digital literacy, further complicate implementation. Additionally, managing group dynamics and ensuring equitable participation in collaborative tasks are identified as potential hurdles.

The review emphasizes the critical role of teacher preparedness in the successful implementation of Web Quests. Many educators require training in digital pedagogy, Web Quest design, and integrating technology into communicative language teaching. Without proper training, the full potential of Web Quest technology may remain untapped.

The findings underscore the transformative potential of Web Quest technology in language education. By promoting active learning, collaboration, and authentic communication, Web Quests address limitations in traditional teaching methods and align with the needs of modern learners. They provide a dynamic platform for integrating technology into the classroom, fostering not only language proficiency but also critical 21st-century skills such as teamwork, problem-solving, and digital literacy.

The implications of these findings are significant for educators, curriculum designers, and policymakers:

- Curriculum Integration: Web Quests should be systematically integrated into language university curricula to support communicative competence development.
- Professional Development: Institutions must invest in teacher training programs to equip educators with the skills needed to design and implement effective Web Quests.
- Equitable Access: Addressing digital disparities is crucial to ensure that all students can benefit from Web Quest-based learning.
- Future Research: Longitudinal studies are needed to evaluate the long-term impact of Web Quest technology on communicative competence and its scalability across diverse educational contexts.

In conclusion, Web Quest technology represents a powerful tool for enhancing foreign language communicative competence, provided that its implementation is carefully planned and supported by adequate resources and training. By addressing existing challenges, educators can harness the full potential of Web Quests to create engaging, effective, and inclusive language learning environments.

Conclusion. This review highlights the transformative potential of Web Quest technology in enhancing foreign language communicative competence among language university students. The findings demonstrate that Web Quests effectively address key aspects of communicative competence, including linguistic accuracy, pragmatic skills, and discourse cohesion, by providing students with structured, inquiry-based tasks rooted in authentic and interactive digital resources. By fostering engagement, motivation, collaboration, and critical thinking, Web Quests offer a dynamic alternative to traditional teaching methods, aligning well with the demands of modern learners and the objectives of communicative language teaching.

Despite these advantages, the review identifies several challenges that must be addressed for successful implementation. These include the need for teacher training, time-intensive task design, managing collaborative dynamics, and overcoming technical barriers such as unequal access to digital tools and varying levels of digital literacy. Addressing these challenges is crucial to maximizing the effectiveness of Web Quest technology in language education.

Suggestions for future research:

- Long-Term Impact Studies: Conduct longitudinal studies to evaluate the sustained impact of Web Quest technology on communicative competence and its transferability to real-world communication contexts.
- Cross-Cultural Applications: Investigate the effectiveness of Web Quests in diverse cultural and linguistic settings to understand their adaptability and scalability.

- Teacher Training Models: Explore innovative professional development programs focused on equipping educators with the skills and resources necessary for effective Web Quest integration.

- Technological Advancements: Examine the integration of emerging technologies, such as artificial intelligence and virtual reality, into Web Quest design to enhance interactivity and learner engagement.

- Equity and Access: Research strategies to mitigate digital disparities and ensure equitable access to Web Quest-based learning for students in under-resourced environments.

- Learner Autonomy: Analyze the role of learner autonomy in Web Quest-based education and identify strategies to balance guided and self-directed learning.

By addressing these research gaps, future studies can contribute to refining Web Quest technology as a powerful and inclusive tool for developing foreign language communicative competence. These efforts will further support educators and institutions in creating effective, innovative, and equitable language learning environments for university students.

References:

1. Dodge, B. (1995). WebQuests: A technique for internet-based learning. The Distance Educator, 1(2), 10–13. https://doi.org/10.1016/j.compedu.2015.04.005

2. Jung, Y., & Lee, H. (2019). The impact of WebQuests on EFL learners' critical thinking and communicative competence. Language Teaching Research, 23(4), 456–472. https://doi.org/10.1177/1362168818780914

3. Pérez, L., & Cárdenas, M. (2021). Task-based learning through WebQuests: Promoting language proficiency and teamwork. Journal of Language and Education, 7(2), 56–72. https://doi.org/10.17323/jle.2021.7.2.56-72

4. Rahimi, M., & Shoghi, M. (2020). Using WebQuest technology to enhance EFL learners' pragmatic skills: Evidence from an experimental study. Computers & Education, 159, 103962. https://doi.org/10.1016/j.compedu.2020.103962

5. Wang, S., & Xie, L. (2021). Bridging the digital divide: A case study on WebQuest implementation in under-resourced language classrooms. Asia-Pacific Education Researcher, 30(3), 211–226. https://doi.org/10.1007/s40299-021-00562-9

UDC 347.1

Aldabergen Gulsim Erkinkyzy,

Suleeva Madina

Master's students
South Kazakhstan Pedagogical University
named after Ozbekali Zhanibekov
(Shymkent, Kazakhstan)

THE IMPACT OF ONLINE LEARNING ON STUDENT MOTIVATION: A QUANTITATIVE ANALYSIS

Abstract: This study investigates the impact of online learning on student motivation across different age groups using a quantitative research approach. The growing integration of digital technologies into education has increased the importance of understanding how online learning environments influence learners' motivation and engagement. Data were collected through an online questionnaire administered to participants from two age groups: 18–22 years and 27 years and above. Respondents evaluated statements related to online learning and motivation using a five-point Likert scale. Descriptive statistical analysis was conducted to compare motivation levels between the groups. The findings revealed that younger learners demonstrated more positive attitudes toward online learning and higher levels of motivation compared to older learners. Participants aged 18–22 reported the highest average scores, indicating greater adaptability to digital learning environments. In contrast, older learners showed lower motivation levels, which may be associated with technological challenges, reduced interaction, and preferences for traditional educational formats. The study highlights the importance of considering age-related differences when designing online learning environments and instructional strategies. The results suggest that enhancing digital literacy, providing technical support, and incorporating interactive learning activities can improve student motivation in online education. Despite limitations related to the small sample size, the study contributes to the understanding of factors influencing motivation in digital learning contexts and provides recommendations for future research and educational practice.

Keywords: online learning, student motivation, digital education, e-learning, age differences, learner engagement, educational technology, self-regulated learning, digital literacy, quantitative analysis.

Introduction. In recent years, education has undergone significant transformation due to the rapid development of digital technologies. The widespread adoption of online learning, particularly following the COVID-19 pandemic, has encouraged educational institutions to integrate digital platforms into teaching and learning processes. As online education becomes increasingly common, understanding its impact on student motivation has become an important area of research.

Student motivation is a key factor influencing academic achievement, engagement, and persistence in learning activities. Motivated learners are more likely to participate actively in the educational process and achieve better learning outcomes [1]. Online learning offers several

advantages, including flexibility, accessibility, and opportunities for independent learning. However, its effectiveness may vary depending on learners' age, digital competence, and learning preferences [2, 3].

Previous studies have demonstrated that online learning environments can support learner autonomy and self-regulated learning, which are closely associated with academic motivation [4]. At the same time, researchers have identified several challenges, including reduced social interaction, technological difficulties, and limited direct communication with instructors, which may negatively affect students' motivation [5, 8].

Age is considered an important factor influencing students' attitudes toward online learning. Younger learners who have grown up in a technology-rich environment tend to adapt more easily to digital learning platforms [6]. In contrast, older learners may experience difficulties adjusting to technology-based instruction and often demonstrate stronger preferences for traditional classroom settings [7].

The purpose of this study is to examine the influence of online learning on student motivation across different age groups. Specifically, the study compares the perceptions and motivation levels of younger and older learners participating in online education.

Literature review. The relationship between online learning and student motivation has been widely investigated in educational research. According to Self-Determination Theory, motivation is influenced by learners' needs for autonomy, competence, and relatedness [12]. Online learning environments can support these needs by providing flexibility and opportunities for independent learning [1].

Hung et al. [2] found that students who demonstrate greater readiness for online learning are more likely to report positive learning experiences and higher levels of motivation. Similarly, Sun et al. [3] identified learner satisfaction, course design, and technological support as significant factors influencing the effectiveness of online learning.

Zimmerman [4] emphasized the importance of self-regulated learning, arguing that students who effectively manage their learning process often demonstrate stronger academic motivation. In online environments, self-regulation becomes particularly important because learners are required to manage their own time and learning activities.

Despite these advantages, several studies have highlighted potential limitations of online learning. Richardson and Swan [8] reported that social presence and interaction significantly influence students' learning satisfaction and motivation. Hartnett [5] further noted that insufficient interaction and low engagement may reduce motivation in online courses.

Research has also suggested that age influences learners' attitudes toward technology-based education. Prensky [6] introduced the concept of "digital natives," referring to younger individuals who are generally more comfortable using digital technologies. Older learners, sometimes described as "digital immigrants," may require additional support when adapting to online learning environments. Moore's Theory of Transactional Distance [7] also suggests that reduced communication and psychological distance may affect learners differently depending on their educational background and experience.

Methods. This study employed a quantitative research design using descriptive and comparative approaches. Data were collected through an online questionnaire consisting of statements related to online learning and student motivation.

A total of eight participants took part in the study. Respondents represented different age groups; however, due to the absence of responses from the 23–26 age category, the analysis focused on two groups: learners aged 18–22 and learners aged 27 and above.

Participants evaluated statements using a five-point Likert scale ranging from strong disagreement to strong agreement. Higher scores indicated more positive attitudes toward online learning and stronger motivation levels.

The collected data were analyzed using descriptive statistical methods. Mean scores were calculated for each age group, and comparisons were made to identify differences in motivation levels. Due to the limited sample size, advanced statistical analyses were not conducted.

Results. The findings revealed noticeable differences between the two age groups regarding their perceptions of online learning.

Participants aged 18–22 demonstrated the most positive attitudes toward online learning, with an average score of 10 positive responses. These results suggest that younger learners are generally more willing to engage with online learning environments and demonstrate higher levels of motivation.

In contrast, participants aged 27 and above obtained a lower mean score of 8.67. This finding indicates that older learners may experience more difficulties when participating in online learning or may prefer traditional learning formats.

The absence of data from the 23–26 age group represents a limitation of the study. This age category often includes university graduates and working professionals whose experiences could provide additional insights into online learning motivation.

Overall, the results suggest that younger learners adapt more successfully to online learning environments than older learners and generally demonstrate stronger motivation toward digital education.

Discussion. The findings of this study are consistent with previous research indicating that younger learners often respond more positively to online learning environments [2, 6]. Their familiarity with digital technologies and confidence in using online platforms may contribute to higher motivation levels.

Older learners, on the other hand, may face challenges related to technology use, reduced interaction, and unfamiliarity with online educational environments [5, 7]. These factors can negatively influence motivation and engagement in digital learning activities.

The study also supports previous findings suggesting that the effectiveness of online learning depends on individual characteristics rather than the learning format alone [1, 11]. Factors such as digital competence, self-regulation skills, and learning preferences play important roles in shaping learners' motivation.

Although online learning offers flexibility and accessibility, educational institutions should consider age-related differences when designing instructional strategies. Providing technical support, clear guidance, and opportunities for interaction may help improve motivation among older learners.

Nevertheless, the study has several limitations. The small sample size and the lack of representation from the 23–26 age group limit the generalizability of the findings. Future research should include larger and more diverse samples and employ additional measurement tools to provide a more comprehensive understanding of online learning motivation.

Practical implications of the study. The findings of this study have important implications for educators, educational institutions, and policymakers involved in the

development of online learning environments. As digital education continues to expand, understanding the factors that influence student motivation becomes increasingly important for ensuring effective learning outcomes.

One of the key findings of this research is that younger learners tend to demonstrate higher levels of motivation in online learning environments than older learners. This suggests that educational institutions should consider age-related differences when designing and implementing online courses. While younger students often adapt quickly to digital technologies, older learners may require additional support, guidance, and training to become comfortable with online learning platforms.

To increase motivation among all learners, instructors should incorporate interactive teaching methods such as discussion forums, collaborative projects, virtual group activities, and multimedia learning materials. These approaches can help create a more engaging learning experience and reduce feelings of isolation that are often associated with online education. Regular feedback from instructors can also contribute to higher levels of motivation by helping students monitor their progress and maintain their engagement throughout the learning process.

Another important consideration is the development of digital literacy skills. Since technological competence influences students' confidence and motivation, educational institutions should provide opportunities for learners to improve their digital skills. Orientation programs, technical support services, and training workshops can help students overcome technological barriers and participate more effectively in online learning activities.

Furthermore, course designers should ensure that online learning environments are user-friendly, accessible, and adaptable to the needs of different learner groups. Clear instructions, well-organized learning materials, and intuitive navigation can reduce cognitive load and improve students' overall learning experience. Flexible learning opportunities may be particularly beneficial for learners who combine education with work or family responsibilities.

Overall, the study demonstrates that motivation in online learning is influenced by multiple factors, including age, technological competence, and individual learning preferences. By addressing these factors, educators can create more inclusive and effective online learning environments that support student engagement, satisfaction, and academic success.

Conclusion. This study examined the impact of online learning on student motivation across different age groups. The findings indicate that younger learners aged 18–22 demonstrate higher levels of motivation and more positive attitudes toward online learning than learners aged 27 and above.

The results emphasize the importance of considering age-related differences when developing online learning environments and instructional practices. While younger learners tend to adapt more easily to digital education, older learners may benefit from additional support and guidance.

Future studies should involve larger participant groups and include all age categories to provide a broader understanding of how online learning influences student motivation.

References:

1. Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. Contemporary Educational Psychology, 61, 101860.

2. Hung, M. L., Chou, C., Chen, C. H., & Own, Z. Y. (2010). Learner readiness for online learning: Scale development and student perceptions. Computers & Education, 55(3), 1080–1090.

3. Sun, P. C., Tsai, R. J., Finger, G., Chen, Y. Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. Computers & Education, 50(4), 1183–1202.

4. Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory Into Practice, 41(2), 64–70.

5. Hartnett, M. (2016). Motivation in online education. Singapore: Springer.

6. Prensky, M. (2001). Digital natives, digital immigrants. On the Horizon, 9(5), 1–6.

7. Moore, M. G. (1993). Theory of transactional distance. In D. Keegan (Ed.), Theoretical Principles of Distance Education (pp. 22–38). London: Routledge.

8. Richardson, J. C., & Swan, K. (2003). Examining social presence in online courses in relation to students' perceived learning and satisfaction. Journal of Asynchronous Learning Networks, 7(1), 68–88.

9. Kauffman, H. (2015). A review of predictive factors of student success in online learning. Research in Learning Technology, 23, 1–13.

10. Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., Wallet, P. A., Fiset, M., & Huang, B. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. Review of Educational Research, 74(3), 379–439.

11. Kahu, E. R. (2013). Framing student engagement in higher education. Studies in Higher Education, 38(5), 758–773.

12. Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. Psychological Inquiry, 11(4), 227–268.

UDC 37.091.33

Tuganbay Yenlik Sydykbaykyzy

Master degree's student
University named after O.Zhanibekov
(Shymkent, Kazakhstan)

LEVERAGING SOCIAL MEDIA CONTENT TO ENHANCE ORAL COMMUNICATION SKILLS IN HIGH SCHOOL ENGLISH LANGUAGE LEARNERS: A REVIEW OF STRATEGIES AND OUTCOMES

Abstract. The integration of social media into English language education has emerged as a promising approach for enhancing oral communication skills among high school students. Social media platforms provide a rich and interactive environment where learners can access authentic language input, engage in meaningful conversations, and develop critical oral proficiency skills such as pronunciation, fluency, and listening comprehension. Unlike traditional teaching methods, which often lack opportunities for real-world practice, social media enables students to immerse themselves in diverse linguistic and cultural contexts, fostering both competence and confidence.

This review aims to evaluate the role of social media content in improving oral communication skills in high school English learners. By systematically analyzing recent literature, the study identifies effective strategies, benefits, and challenges associated with integrating social media into language education. The methodology involved a systematic review of academic sources published in recent years, focusing on studies that explored the relationship between social media use and oral communication skill development. The findings reveal that social media enhances student engagement and motivation, provides exposure to natural speech patterns, and facilitates interactive practice through features like live chats, comments, and video responses.

However, the review also highlights challenges, including disparities in digital access, teacher preparedness, and the potential for distractions. It underscores the importance of professional development for educators, structured lesson planning, and equitable access to technology to ensure successful implementation. Despite these challenges, the review concludes that social media offers significant opportunities to bridge the gap between theoretical learning and practical application in language education.

The significance of this study lies in its ability to inform educators and policymakers about the potential of social media as a transformative tool in language learning. By integrating these platforms effectively, educators can create engaging and dynamic learning environments that support the development of oral communication skills. The review also identifies areas for further research, including the long-term impact of social media on language proficiency and strategies to address access and equity issues. Ultimately, this study provides a foundation for leveraging social media to meet the evolving needs of high school English learners and prepare them for real-world communication challenges.

Keywords: oral proficiency, digital education, authentic communication, interactive learning, multimodal engagement, language fluency, cultural competence

Introduction. In an era of digital connectivity, social media has emerged as a powerful tool that transcends traditional boundaries of communication and education. For high school students learning English as a second language, social media platforms offer access to authentic language content, enabling immersive exposure to real-life communication. This is particularly crucial for developing oral proficiency, a skill that often requires interactive practice and cultural context, which traditional classroom resources may fail to provide. The integration of social media content into English language education thus represents a promising approach to bridging the gap between theoretical knowledge and practical language use.

The relevance of this topic lies in the increasing prevalence of social media in the daily lives of students. Platforms such as YouTube, TikTok, Instagram, and others provide rich, engaging content that can serve as valuable resources for language learning. These platforms not only expose students to diverse accents, expressions, and contexts but also foster interactive communication through comments, live discussions, and other participatory features. Despite the potential of these tools, their effective integration into English language education remains underexplored, particularly in the context of oral communication development for high school students.

The primary objective of this study is to evaluate how social media content can be effectively leveraged to enhance the oral communication skills of high school students learning English. This review aims to analyze existing strategies for integrating social media into language learning, identify the benefits and challenges associated with its use, and explore its implications for teaching practices and curriculum design. By addressing these objectives, the study seeks to provide a comprehensive understanding of how social media can be transformed from a casual entertainment medium into a powerful educational resource for developing oral proficiency in English language learners.

Theoretical background. The integration of social media into education, particularly in language learning, has been extensively studied in recent years, reflecting its growing significance as a tool for enhancing oral communication skills. This section reviews existing literature on the role of social media in foreign language education, the theoretical frameworks underpinning its use, and its specific contributions to the development of oral proficiency.

Social media platforms such as YouTube, TikTok, Instagram, and Facebook provide authentic and diverse linguistic content that mirrors real-world language use. Studies have shown that these platforms facilitate exposure to natural speech patterns, colloquial expressions, and cultural nuances, which are essential for developing oral communication skills. For instance, video-based platforms allow learners to observe and imitate pronunciation, intonation, and body language, creating a virtual immersion experience [1].

Interactive features of social media, such as live chats, comment sections, and collaborative projects, also encourage active language use. These opportunities for real-time communication help students practice speaking and listening skills in authentic, meaningful contexts. Furthermore, the multimedia nature of social media accommodates various learning styles, enhancing students' engagement and motivation [2].

Oral communication encompasses several interconnected skills, including pronunciation, fluency, listening comprehension, and interactive competence. Traditional language instruction often falls short in providing sufficient practice for these skills due to limited classroom time and the lack of authentic communication opportunities. Social media addresses these gaps by

offering dynamic environments where learners can practice speaking and listening in real-life scenarios [2].

Research has highlighted that exposure to authentic content on social media helps learners improve their pragmatic competence, or the ability to use language appropriately in specific social contexts. For example, participating in live discussions or engaging with influencers who speak the target language allows students to learn how to express opinions, respond to questions, and maintain conversations naturally [3].

The integration of social media into language learning is grounded in several educational theories:

- Constructivism: This theory emphasizes learning through interaction and experience. Social media provides a constructivist environment where students actively engage with content, interact with peers, and construct their understanding of language use.

- Communicative Language Teaching (CLT): CLT focuses on developing learners' ability to communicate effectively. Social media aligns with this approach by offering contexts for meaningful communication and interaction.

- Vygotsky's Sociocultural Theory: According to this theory, learning occurs through social interaction. Social media enables learners to engage with a global community, fostering collaborative learning and peer feedback.

Despite its benefits, the integration of social media into language education presents challenges. Teachers may lack the training to effectively incorporate these platforms into their curriculum, and students may face distractions or misuse the tools for non-educational purposes. Additionally, disparities in access to technology and reliable internet connections can limit the effectiveness of social media as a learning resource [4].

The findings from existing literature suggest that social media has significant potential to enhance oral communication skills, but its use must be carefully planned and integrated into the curriculum. Structured activities, such as guided discussions, video assignments, and peer feedback, can help maximize the educational benefits of social media while minimizing its drawbacks. Furthermore, educators must develop strategies to align social media activities with specific learning objectives and provide clear guidelines to students [5].

By synthesizing these insights, this review establishes a theoretical foundation for understanding how social media content can be effectively utilized to develop oral communication skills in high school English language learners. It also highlights the need for further research to explore the long-term impact and scalability of these approaches in diverse educational contexts.

Methodology. This review employs a systematic approach to literature selection and thematic analysis to investigate the integration of social media content for improving oral communication skills in high school English language learners. The methodology ensures a comprehensive and objective analysis of recent and relevant studies. The process of literature selection involved multiple stages. First, a database search was conducted using academic platforms such as Google Scholar, ERIC, Scopus, and PubMed to identify peer-reviewed journal articles, conference proceedings, and book chapters published between recent years. Keywords such as "social media integration," "oral communication skills," "high school English learners," "digital tools in education," and "language learning through social media" were used, combined with Boolean operators to refine the search results.

Studies were included if they specifically focused on the use of social media in foreign language learning, examined its impact on oral communication skills, targeted high school students or similar age groups, and provided empirical data or theoretical insights. Articles were excluded if they primarily addressed higher education, were written in languages other than English, or lacked methodological rigor. After the initial search, duplicate records were removed, and the remaining studies were shortlisted based on their titles and abstracts. The full texts of these studies were thoroughly reviewed, and the articles that met all inclusion criteria were selected for the final analysis.

The thematic analysis of the selected studies was carried out to identify recurring patterns, themes, and research gaps. This process involved systematically coding the key findings and concepts from each study. Codes such as “engagement,” “authentic content,” “peer interaction,” “oral fluency,” and “teacher challenges” were used to categorize the data. These codes were then grouped into broader themes, including the benefits of social media for oral proficiency, the challenges of integrating digital tools in the classroom, and strategies for effective curriculum design. Themes were reviewed and refined to ensure they accurately represented the findings from the literature.

This methodology provided a robust framework for synthesizing diverse perspectives and insights into the use of social media content for enhancing oral communication skills. It also enabled the identification of best practices, limitations, and areas requiring further exploration, thereby contributing to a deeper understanding of this emerging field in language education.

Findings and Discussion. The findings of this review reveal significant insights into the potential of integrating social media content into high school English language education, particularly for improving oral communication skills. The results are presented in key thematic areas, followed by a discussion of their implications.

The review identifies that social media platforms provide unique opportunities for students to practice oral communication in authentic, real-world contexts. Platforms such as YouTube, Instagram, and TikTok expose learners to diverse linguistic inputs, including various accents, speech patterns, and cultural nuances. This exposure improves listening comprehension and pronunciation. Moreover, interactive features such as live discussions, comment threads, and video responses allow students to actively engage in speaking and conversational practice, fostering fluency and confidence.

The findings emphasize that the use of multimedia and visual aids available on social media strengthens memory retention and supports multimodal learning. These resources provide a rich environment for oral skill development, bridging the gap between classroom instruction and real-life communication.

Social media platforms are inherently engaging for high school students, making them a powerful tool for motivating learners. The studies reviewed highlight that integrating familiar and enjoyable tools into the learning process enhances students’ willingness to participate in speaking activities. By leveraging content they encounter daily, such as influencer videos or trending discussions, students feel more connected to the material, which boosts their enthusiasm for language practice.

Despite the benefits, the review identifies significant challenges associated with integrating social media into language education. Teachers face difficulties in curating appropriate content that aligns with curriculum goals and ensures language accuracy. Additionally, disparities in access to digital tools and reliable internet connections can create

inequities among students. Concerns about the potential for distractions and misuse of social media further complicate its implementation.

The findings underscore the importance of teacher training and support for the successful integration of social media content. Many educators lack the necessary digital skills to effectively incorporate these platforms into their teaching. Professional development programs that focus on digital pedagogy and content curation are critical for overcoming this barrier.

The findings have significant implications for educational practices and curriculum design. By integrating social media content, educators can create a more dynamic and engaging language learning environment that supports the development of oral communication skills. However, to maximize its potential, structured guidelines and strategies must be developed to address challenges such as access disparities and teacher preparedness.

Moreover, educational institutions and policymakers should invest in infrastructure and training to ensure equitable access to technology and equip teachers with the skills to effectively use social media as a learning tool. Future research should explore scalable models for integrating social media into language education and investigate its long-term impact on oral proficiency and overall language competence.

In conclusion, the review highlights that while social media offers transformative opportunities for enhancing oral communication skills in high school English language learners, its successful integration requires careful planning, resources, and ongoing support. The insights gained provide a foundation for further exploration and practical application in educational contexts.

Table 1- Key aspects of using social media for enhancing oral communication skills in language learning

Aspect	Benefits	Challenges	Strategies
Engagement	Social media platforms increase student motivation and participation through engaging content.	Students may become distracted by non-educational content.	Establish clear usage guidelines and focus on task-based activities.
Authentic Input	Provides exposure to real-world language, accents, and cultural nuances.	Difficulty in finding content that aligns with curriculum goals.	Curate content in advance that meets educational standards.
Practice Opportunities	Interactive features like live chats and comment sections enhance speaking and listening practice.	Unequal access to devices and internet among students.	Ensure equitable access to technology through school resources.
Skill Development	Enhances pronunciation, fluency, and pragmatic competence.	Teachers may lack skills to integrate digital tools effectively.	Offer professional development programs focused on digital pedagogy.
Multimodal Learning	Combines video, audio, and text to cater to different learning styles.	Overwhelming variety of content can make it hard for students to focus.	Design structured lesson plans integrating specific social media content with clear objectives.
Student Autonomy	Encourages independent exploration and learning outside the classroom.	Students might misuse the platform for non-educational purposes.	Use monitoring tools and provide clear expectations for online activities.

Cultural Exposure	Introduces students to authentic cultural contexts and idiomatic expressions.	May expose students to inappropriate or misleading content.	Select age-appropriate and culturally relevant materials in advance.

Conclusion. This review highlights the transformative potential of integrating social media content into English language education for high school students, specifically for improving oral communication skills. Social media platforms offer authentic linguistic input, engaging learning environments, and opportunities for interactive practice, which are essential for developing pronunciation, fluency, and cultural competence. The review also underscores the motivational power of social media, as its familiarity and multimedia features increase student engagement and enthusiasm for language learning.

However, the findings also reveal several challenges, including teacher preparedness, ensuring appropriate and curriculum-aligned content, and addressing disparities in access to digital tools. Effective integration of social media into education requires thoughtful planning, structured implementation, and teacher training programs that equip educators with the necessary skills to use these tools effectively. Additionally, clear guidelines and monitoring are essential to prevent distractions and misuse by students.

Suggestions for future research:

- Long-Term Impact Studies: Investigate the sustained impact of social media integration on oral communication skills and overall language proficiency in diverse educational settings.
- Scalable Models: Explore scalable strategies for implementing social media-based learning in different school environments, particularly in resource-constrained contexts.
- Teacher Training Programs: Develop and evaluate training programs that focus on equipping educators with the skills to curate, design, and integrate social media content into their teaching.
- Equity and Access: Examine the effectiveness of policies and practices aimed at ensuring equitable access to digital tools and platforms, addressing the digital divide in education.
- Interdisciplinary Approaches: Study the intersection of social media use in language education with emerging technologies, such as artificial intelligence and virtual reality, to explore their combined potential for enhancing oral communication.

By addressing these areas, future research can provide deeper insights and practical solutions for leveraging social media to create more effective and inclusive language learning environments.

References:

1. Richards, J. C., & Schmidt, R. (2020). Longman Dictionary of Language Teaching and Applied Linguistics. Routledge.

2. Godwin-Jones, R. (2019). Using social media for language learning: Emerging perspectives and practices. Language Learning & Technology, 23(1), 10–21.

3. Wang, S., & Vásquez, C. (2020). Web 2.0 and second language learning: What does the research say? CALICO Journal, 37(2), 137–159.

4. Sykes, J. M., & Reinhardt, J. (2021). Language at Play: Digital Games in Second and Foreign Language Teaching and Learning. Pearson Education.

5. Thorne, S. L., & Lantolf, J. P. (2020). Social networking tools for second language acquisition. Annual Review of Applied Linguistics, 40, 147–163.

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

UDC 339.13

Dinara Abdesh

11th grade student
Supervisor: Serzhan Balgyn
Nazarbayev Intellectual School
(Astana, Kazakhstan)

TWO LEVERS, ONE DECISION: HOW PRICE DISCOUNTS AND NON-FINANCIAL SERVICE INTERACT TO SHAPE PURCHASE DECISIONS AND STORE REVENUE IN KAZAKHSTANI APPAREL RETAIL

Annotation: Retailers in Kazakhstan compete on two levers at once: the price they put on a tag, and everything around that price, how a shop assistant talks to a customer, how the store feels, how the product is shown. Economics usually studies these separately. Price discounts belong to demand theory and the psychology of deals; service belongs to a different literature on service quality. This paper argues they should be read together, because in a real clothing store they act on the same customer at the same moment, and sometimes they pull in opposite directions. I review the literature on discount psychology, reference-price effects, and retail service quality, and I use it to build a simple framework for an emerging-market apparel setting like the one I worked in over a summer. To make the demand logic concrete I include a clearly hypothetical, textbook demand schedule and use it to separate two ideas that are easy to confuse: the price that maximizes revenue and the price that maximizes profit. The contribution is conceptual, not empirical: I do not have store data, and I am careful to say where my own observations are anecdotes rather than evidence. What I offer is a structured way to think about the interaction of discounts and non-financial service in Kazakhstani retail, an interaction the existing studies, which mostly look at one lever or the other, tend to leave on the side. I close by laying out the empirical study this framework asks for, which is the natural next step.

Keywords: price discounts, price elasticity, reference price, service quality, consumer behavior, retail, Kazakhstan

1. Introduction

Walk into almost any clothing store in Astana and you see two things happening. There is a price, often with a discount sticker on it. And there is a person, a sales consultant, who decides how much attention to give you. Most of the time we treat these as two different worlds. Price is "economics." Service is "soft." But for the customer holding a shirt and deciding whether to buy it, they arrive together.

I became interested in this because I spent the summer of 2025 working as a sales consultant at Big Cotton, a clothing store in Kazakhstan, in June, July, and August. Big Cotton is known for being affordable, you can find clothes, shoes, and accessories there for under 10,000 tenge, sometimes much less. My pay was a fixed wage plus a 2% bonus on what I sold,

so I had a personal reason to notice what actually moved a sale. What I noticed did not fit the simple story that "lower price means more sales." Some days a discount did almost nothing. Some days a customer who was clearly leaving stayed and bought because I asked the right question, helped them choose, and removed the small doubts that were stopping them. I want to be honest about what that is: it is my own experience, not a dataset. It is what made me ask the question, not evidence for the answer.

The question this paper takes up is: how do price discounts and non-financial service interact to shape a customer's purchase decision and, through that, a store's revenue, in an emerging-market clothing context like Kazakhstan's?

Retail is a good place to ask this. It reacts fast, the margins are thin, and a manager genuinely has to choose between cutting the price and investing in the people and the experience on the floor. The economics of discounts is well studied, and so is service quality, but they are usually studied in separate rooms. This paper tries to put them in the same room. It is a conceptual paper. I do not estimate anything from real store records, I do not have them yet, and I will not pretend otherwise. Instead I read the existing research, build a framework from it, and use a transparent, hypothetical example to keep the economic logic honest. At the end I describe the empirical study that this framework points to, which is the work I would do next if I can get real data from the store.

The paper is organized as follows. Section 2 reviews three strands of literature: the psychology of discounts, reference-price and behavioral pricing, and retail service quality, and names the gap I am trying to address. Section 3 lays out a conceptual framework and a clearly-labeled illustrative demand example that separates revenue from profit. Section 4 discusses what this means for Kazakhstani retailers. Section 5 states the limitations plainly. Section 6 concludes and describes the empirical next step. Appendix A gives the full worked example behind the figures.

2. Literature review

2.1 The psychology of discounts

A discount is not only a smaller number. It is read by the customer as a gain, and people respond to gains and losses in a way that is not symmetric or fully rational. Kahneman and Tversky's prospect theory showed that we judge outcomes against a reference point and that losses loom larger than equivalent gains (Kahneman & Tversky, 1979). A price marked down from a "before" price is experienced relative to that anchor, so the same final price feels better when it arrives as a discount than when it is simply the price. This is the mechanism under a lot of everyday retail behavior: urgency from limited-time offers, the pull of "everyone is buying it," the relief of feeling you saved money.

There is also direct evidence on how discounts work as a sales tool. Blattberg, Briesch, and Fox (1995) synthesized a large body of empirical findings on price promotions and drew out regularities, for example, that promotions reliably lift short-term sales, but with effects that depend on the category and that do not simply scale up forever. Della Bitta, Monroe, and McGinnis (1981) studied how the depth and presentation of a discount change what customers perceive, which matters because a 20% sticker and a 60% sticker are not just two points on a line; they are read differently. More recently, Büyükdağ, Soysal, and Kitapçı (2020) ran an experiment on specific discount patterns and showed that how a price promotion is framed changes perceived price attractiveness and purchase intention. The level of method in that paper,

a controlled experiment with measured intention, is the standard the empirical version of my own question would have to meet.

The honest summary of this strand: discounts work, but through perception, and with limits.

2.2 Reference price and the discount paradox

If discounts only ever helped, no manager would hesitate. But they can backfire, and the reason is again about reference points. Putler (1992) built reference-price effects directly into a model of consumer choice, formalizing the idea that today's purchase is judged against a remembered or expected price. Once customers learn to expect discounts, the discounted price becomes the reference, and the next full price feels like a loss. A deep or constant discount can also act as a signal: if it is this cheap, maybe it is not very good. This is the "discount paradox", past some point, cutting the price more can lower what the customer is willing to do, especially for a brand whose buyers care about style or image.

Grewal, Krishnan, Baker, and Borin (1998) sit closest to my own question. They studied how store name, brand name, and price discounts together shape consumers' evaluations and purchase intentions, that is, they already treat price and a non-price factor (store/brand reputation) as acting jointly, not separately. Their work is the nearest neighbor to what I want to do, and it is also where I can see the opening: their non-price factor is reputation, set before the customer arrives, while the lever I watched all summer was interaction at the moment of sale, what the person on the floor does. That is a different non-financial service, and it is much less studied in this combined way.

2.3 Retail service quality

The second lever has its own mature literature. Parasuraman, Zeithaml, and Berry (1988) introduced SERVQUAL, a multi-item scale for measuring perceived service quality across dimensions like reliability, responsiveness, and assurance. The key point for me is that service quality is measurable, there is an accepted instrument, even though in the original paper I read, and in most casual discussion, it is treated as a vague "be nice to customers." If the effect of service on sales is ever going to be tested rather than asserted, it has to be measured this way, not inferred from how much I personally happened to earn on a given day.

On the demand side, there is also real work on estimating apparel demand and elasticity from actual data. Jones (2002) estimated the economic determinants of clothing consumption in the UK, which is the kind of real estimation a credible empirical paper on my topic would resemble. Guadagni and Little (1983) is the classic example of modeling store-level choice from real scanner data, the gold standard for "demand from real records." I cite these not because this paper does that, but to be clear about the distance between a conceptual framework (what this is) and an estimated model (what the next step would be).

2.4 Reading the literature against my question

It helps to set the nine sources side by side and ask, for each, what it gives the question and where it stops short of it. Table 1 does this. The pattern that comes out is the gap: each strand is strong on its own lever, the one paper that combines a price and a non-price lever (Grewal et al., 1998) uses a non-price lever fixed before the customer arrives, and almost none of it is set in an emerging-market value-apparel context.

Table 1. Literature synthesis: each source, its core finding, and how it relates to the gap this paper addresses.

Source	Strand	Core finding
Kahneman & Tversky (1979)	Behavioral foundation	Choices are judged against a reference point; losses loom larger than equal gains
Blattberg, Briesch & Fox (1995)	Discount psychology	Promotions reliably lift short-term sales, but effects are category-dependent and do not scale without limit
Della Bitta, Monroe & McGinnis (1981)	Discount psychology	Discount depth and presentation change perceived value, not just the final price
Büyükdağ, Soysal & Kitapçı (2020)	Discount psychology (method bar)	An experiment shows discount pattern changes perceived price attractiveness and purchase intention
Putler (1992)	Reference price	Formal reference-price effects in consumer choice; asymmetric response around the reference
Grewal, Krishnan, Baker & Borin (1998)	Price × non-price (nearest neighbor)	Store name, brand name, and discount jointly shape evaluations and purchase intentions
Parasuraman, Zeithaml & Berry (1988)	Service quality	SERVQUAL: service quality is measurable on a validated multi-item scale
Jones (2002)	Demand estimation	Estimates the economic determinants of UK clothing consumption from real data
Guadagni & Little (1983)	Demand estimation	Logit brand-choice model calibrated on real scanner data

2.5 The gap

Three things stand out from reading across these strands. First, discount psychology and service quality are each well developed, but mostly in separate literatures. Second, the studies that do combine price with a non-price factor, Grewal et al. (1998) being the clearest, use a non-price factor that is fixed before the customer walks in (store/brand reputation), not the live, in-store service interaction that a salesperson actually controls. Third, very little of this is set in an emerging-market, value-oriented apparel context like Kazakhstan's, where price sensitivity is high and the in-store consultation is a normal part of buying clothes.

So the gap I am addressing is narrow and, I think, real: the interaction between price discounts and in-store non-financial service, in an emerging-market apparel setting, treated as

two levers acting on the same purchase decision at the same time. This is a conceptual contribution, I am organizing the existing ideas into a framework for this specific case and pointing to the test it implies. I am not claiming to have measured the interaction.

3. Conceptual framework

3.1 Two levers, one decision

The framework is simple to state. A customer's decision to buy is shaped by (1) the price, read through the lens of discounts and reference points, and (2) the non-financial service around the price, which lowers uncertainty and builds enough trust to close the sale. These are not additive and independent. They interact:

When price sensitivity is high (an elastic, value-shopping customer), a discount does a lot of the work, and service mainly removes the last small doubts.

When the customer is uncertain, unsure about fit, quality, or whether they actually need the item, service can do what a discount cannot, by answering questions and reducing perceived risk.

And the two can pull against each other: a very deep discount can trigger the discount-paradox signal ("is this cheap because it is bad?"), and good service can be exactly what counteracts that signal, by re-establishing the product's value through attention and explanation.

That last case is the interesting one, and it is the reason to study the two levers together rather than apart. Figure 2 draws the framework: the two levers on the left act through two mediators, perceived value and perceived risk, on the probability of purchase, and through it on revenue, while the dashed loop shows the discount paradox feeding back on perceived value and service counteracting it. The diagram is a framework, not a result; nothing in it is measured.

Figure 2. Conceptual framework: two levers acting on one purchase decision (a framework, not a result)

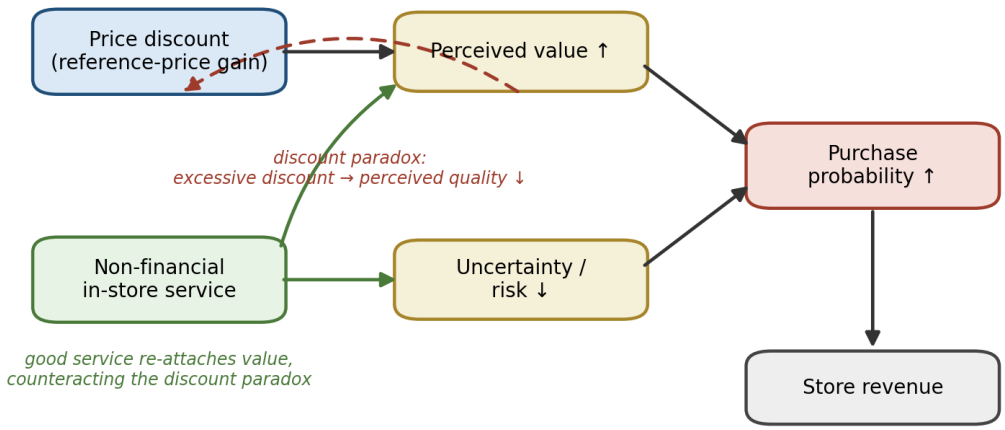


Figure 2. Conceptual framework: price discount and non-financial service act through perceived value and perceived risk on purchase probability and revenue, with the discount-paradox feedback loop. This is a framework, not an estimated result.

The behavioral factors and the service responses to them can be organized as in Table 2, adapted from the structure in the original project and re-grounded in the literature above.

Table 2. Behavioral factors at the point of sale, the non-financial service response to each, and the likely economic effect.

Behavioral factor	Non-financial service response	Likely effect on the purchase	Grounded in
Uncertainty when choosing	Consultation, comparing options, showing the product	Higher probability of purchase; larger basket	Parasuraman et al. (1988); Grewal et al. (1998)
Wish to avoid risk	Guarantees, fitting help, after-sale support	More confidence in the choice; fewer abandoned sales	Parasuraman et al. (1988)
Wanting to feel looked after	Polite, personal attention; a comfortable atmosphere	Loyalty; repeat visits	Parasuraman et al. (1988)
Reading the price as a deal	Honest framing of the discount against a fair reference	Higher perceived value without triggering the paradox	Kahneman & Tversky (1979); Putler (1992); Della Bitta et al. (1981)
Reacting to a very deep discount	Service that re-attaches value (explanation, attention)	Protects perceived quality; offsets the discount paradox	Putler (1992); Grewal et al. (1998)

Table 2 is a conceptual map, not a result. Each row is a hypothesis the literature makes plausible and that the future empirical study would test.

3.2 An illustrative demand example (hypothetical, not estimated)

To keep the economic logic concrete and honest, it helps to work through a demand curve. The example below is entirely hypothetical. It is a textbook demand schedule chosen to be simple and internally consistent; it is not estimated from Big Cotton or any real data, and the numbers should not be read as findings about any real store. Its only job is to make the relationship between price, quantity, revenue, and elasticity transparent, and to separate two ideas the original version of this work confused.

Take a simple linear demand schedule for one apparel item:

$$Q = 700 - 0.05 \times P$$

where P is the price in tenge and Q is units sold over some period. This says: when the price rises, quantity falls, in a straight line. Plotting it together with revenue gives Figure 1, and reading values off it gives Table 3.

Figure 1. Illustrative linear demand and revenue (hypothetical — not estimated from data)

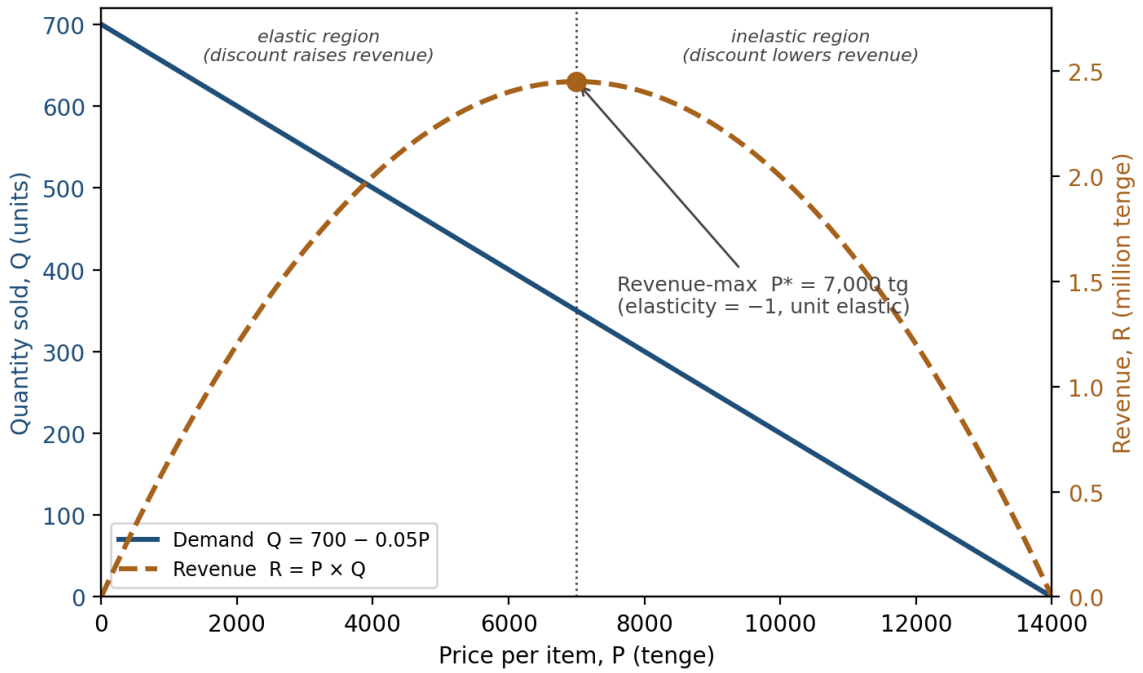


Figure 1. Illustrative linear demand curve (Q = 700 − 0.05P) and the revenue curve R = P × Q, with the revenue-maximizing price marked at P* = 7,000 tenge (unit elasticity). Hypothetical illustration, not estimated from data.

Table 3. Illustrative (hypothetical) demand schedule, with revenue and point elasticity. The calculations are shown in Appendix A

Price P (tg)	**Quantity Q (units)**	**Revenue R = P × Q (tg)**	**Point elasticity**
10,000	200	2,000,000	−2.50
8,000	300	2,400,000	−1.33
7,000	350	2,450,000	−1.00
6,000	400	2,400,000	−0.75
4,000	500	2,000,000	−0.40
2,000	600	1,200,000	−0.17

Two things in this table are worth slowing down on.

Revenue peaks where demand is unit-elastic. For a linear demand curve Q = a − bP, revenue R = P(a − bP) is maximized at P* = a/(2b). With a = 700 and b = 0.05 that is P* = 7,000 tenge, where Q = 350, revenue = 2,450,000 tenge, and the point elasticity is exactly −1. Above 7,000 demand is elastic (elasticity below −1): cutting the price raises revenue, because the extra units more than make up for the lower margin per unit. Below 7,000 demand is inelastic: cutting the price lowers revenue, because the extra units no longer compensate. This is the real shape of the discount story, discounts help up to a point and then stop helping, and it falls straight out of the curve, with no fitting and no invented data. Figure 1 marks that turning point.

Revenue is not profit. This is the single most important correction to make. Revenue is price times quantity, R = P × Q. It says nothing about cost. Profit is (price − unit cost) × quantity. They are maximized at different prices. If the store buys each item for a unit cost c, profit is maximized at P = (a + b·c)/(2b), which is higher than the revenue-maximizing price whenever the cost is positive. With this curve:

unit cost 3,000 tg → profit-max price = 8,500 tg (revenue-max was 7,000);

unit cost 4,000 tg → profit-max price = 9,000 tg;
unit cost 5,000 tg → profit-max price = 9,500 tg.

Figure 3 draws this for a unit cost of 4,000 tenge: the profit curve peaks to the right of the revenue curve. So a manager who finds the "revenue-maximizing discount" and stops has not found the profit-maximizing discount, the profit-maximizing price sits higher, because every unit sold at a deeper discount eats into margin. You cannot read a "best discount for profit" off a revenue calculation, because a revenue calculation never sees the costs.

Figure 3. Revenue is not profit: the profit-max price sits above the revenue-max price (hypothetical illustration — not estimated from data)

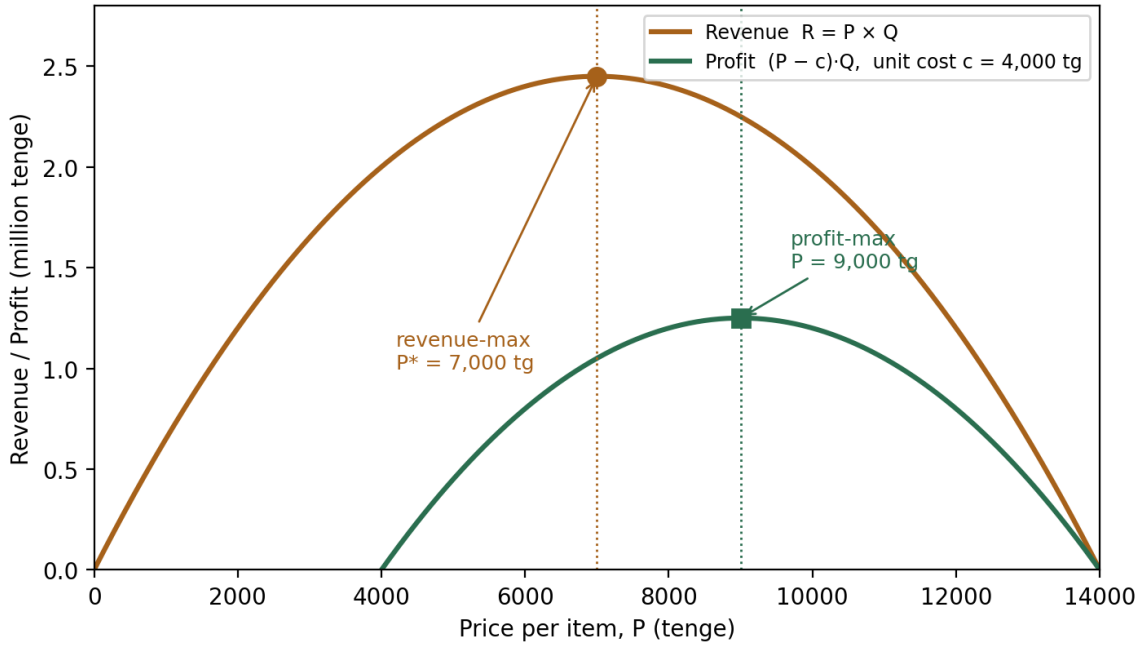


Figure 3. Revenue versus profit for the same illustrative demand curve at unit cost c = 4,000 tenge: the profit-maximizing price (9,000 tg) sits above the revenue-maximizing price (7,000 tg). Hypothetical illustration, not estimated from data.

The point of the example is not the specific numbers, they are made up on purpose. The point is the structure: there is a revenue-maximizing price, it sits where demand turns from elastic to inelastic, and it is not the same as the profit-maximizing price. Any claim about an "optimal discount" has to say which of the two it means. The full derivation, and the script that reproduces every number in Table 3 and both figures, is in Appendix A.

4. Discussion: implications for Kazakhstani retailers

If the framework is right, a few practical things follow for a value-oriented clothing retailer in Kazakhstan.

First, discounts are a real but bounded tool. The elastic region is where they earn their keep; pushing into the inelastic region trades away revenue for volume that does not pay for itself, and risks the discount-paradox signal on a brand whose customers still care about looking good. A store does not need the deepest discount; it needs the discount that sits in the elastic region and stops there.

Second, non-financial service is not a soft extra, it is the lever that works precisely where discounts fail. When a customer is uncertain rather than price-driven, attention, honest comparison, and risk reduction can close a sale that no sticker would. And good service is plausibly what protects a brand from the trust-damage of a deep discount, by re-attaching value

to the product through the interaction. In a market where the in-store consultation is normal and price competition is fierce, this is exactly the margin a store can defend.

Third, and this is the reason to study them together, the two levers should be set jointly, not in separate meetings. A pricing decision that ignores the service experience, or a service-training decision that ignores where the price sits on the demand curve, is optimizing half the problem. The cautious version of the recommendation is: use discounts for short-term demand inside the elastic region, and invest in service as the longer-term lever that holds value and brings customers back. I keep this cautious on purpose, because at this stage it follows from a framework and from theory, not from a measured effect in a real store.

5. Limitations

I want to be clear about what this paper is and is not.

It is conceptual. There is no dataset, no sample size, and no statistical test. The demand example in Section 3.2, and Figures 1 and 3, are hypothetical and labeled as such; they illustrate economic logic and should not be read as results about any real store, including Big Cotton. Figure 2 is a framework, not an estimate.

My own retail experience is an anecdote, not evidence. I noticed that my daily sales bonus moved a lot from day to day, and that it tended to be higher on days when I gave fuller consultations. But that bonus is simply 2% of what I sold, so using it to "show" that service drives sales would be circular, it is revenue measuring itself. And the swings could just as easily come from the day of the week, foot traffic, the weather, what was in stock, or which shift I worked. So I have deliberately demoted that observation to what it really is: the thing that made me curious, not a finding. The original version of this project treated it as evidence, and that was a mistake I am correcting here.

The framework also has scope limits. It is built for value-oriented apparel in an emerging market; it would not transfer cleanly to luxury goods, to groceries, or to pure online retail, where the service interaction is different or absent.

Finally, the literature I lean on is mostly from developed markets. Whether the same patterns hold in Kazakhstan is an open question, which is part of why the empirical step matters.

6. Conclusion and next step

Price discounts and non-financial service are usually studied apart, but they act on the same customer at the same moment, and sometimes against each other. This paper has argued for reading them together in an emerging-market apparel context like Kazakhstan's, reviewed the literature on discount psychology, reference price, and service quality, and used a transparent hypothetical demand example to keep the economics honest, in particular, to separate the revenue-maximizing price from the profit-maximizing price, which are not the same.

The contribution is conceptual: a framework and a clearly-stated gap, not a measured effect. The honest next step is the empirical study this framework asks for. With real store records from a shop like Big Cotton, daily price and units sold for one apparel category over several weeks, one could estimate an actual demand curve and price elasticity for that category, in the style of Jones (2002). With a real, separate measure of service quality (a SERVQUAL-type instrument, or a structured record of consultation intensity that is not derived from sales), one could begin to test whether service and discounts interact the way the framework predicts. That is the study I would like to do, and it is the version of this work that would move it from a conceptual argument to evidence. This paper is the groundwork for it.

Appendix A. The worked example behind the figures

This appendix gives the full derivation behind Table 3, Figure 1, and Figure 3, so every quantitative claim in Section 3.2 is checkable by hand. All of it is hypothetical, a textbook schedule, not data.

The demand schedule. I use the linear demand curve

$$Q(P) = a - bP, \text{ with } a = 700 \text{ and } b = 0.05,$$

so that two convenient anchor points hold exactly: at $P = 10{,}000$, $Q = 200$; at $P = 4{,}000$, $Q = 500$. The slope is $b = (500 - 200) / (10{,}000 - 4{,}000) = 300 / 6{,}000 = 0.05$, and the intercept is $a = 700$.

Revenue. Revenue is $R(P) = P \cdot Q(P) = P(a - bP) = aP - bP^2$. This is a downward parabola in P. Setting the derivative to zero,

$$dR/dP = a - 2bP = 0 \Longrightarrow P^* = a / (2b) = 700 / 0.10 = 7{,}000 \text{ tenge.}$$

At $P^* = 7{,}000$: $Q = 700 - 0.05 \times 7{,}000 = 350$ units, and $R = 7{,}000 \times 350 = 2{,}450{,}000$ tenge. This is the peak of the revenue curve in Figure 1.

Elasticity. The point own-price elasticity of demand is

$$E(P) = (dQ/dP) \cdot (P / Q) = -b \cdot P / (a - bP).$$

At $P^* = 7{,}000$, $E = -0.05 \times 7{,}000 / 350 = -1.00$, i.e. demand is unit-elastic exactly at the revenue-maximizing price, which is the general result for any linear demand curve. Above 7,000, $|E| > 1$ (elastic); below 7,000, $|E| < 1$ (inelastic). The elasticity column of Table 3 is computed this way at each price.

Profit and why its peak is higher. With a constant unit (purchase) cost c, profit is

$$\pi(P) = (P - c) \cdot Q(P) = (P - c)(a - bP).$$

Maximizing, $d\pi/dP = a - 2bP + bc = 0$, which gives

$$P_{profit} = (a + bc) / (2b) = P^* + c/2.$$

Because $c > 0$, the profit-maximizing price always exceeds the revenue-maximizing price by exactly $c/2$. For $c = 3{,}000$, $4{,}000$, $5{,}000$ tenge this gives $P_{profit} = 8{,}500$, $9{,}000$, $9{,}500$ tenge, the numbers in Section 3.2, and the $c = 4{,}000$ case drawn in Figure 3.

Reproducibility. Every number above, the full Table 3, and both figures are produced by two short scripts committed with this paper: analysis/illustrative_demand.py (the schedule, revenue, elasticity, and the revenue-vs-profit prices) and analysis/make_figures.py (Figures 1–3).

Running it prints Table 3, the revenue-max price (7,000 tg, unit-elastic), and the three profit-max prices (8,500 / 9,000 / 9,500 tg) used above.

References:

1. Blattberg R. C., Briesch R., & Fox E. J. How promotions work. Marketing Science, 14(3, Suppl.), G122–G132, 1995. https://doi.org/10.1287/mksc.14.3.g122

2. Büyükdağ N., Soysal A. N., & Kitapçı O. The effect of specific discount pattern in terms of price promotions on perceived price attractiveness and purchase intention: An experimental research. Journal of Retailing and Consumer Services, 55, 102112, 2020. https://doi.org/10.1016/j.jretconser.2020.102112

3. Della Bitta A. J., Monroe K. B., & McGinnis J. M. Consumer perceptions of comparative price advertisements. Journal of Marketing Research, 18(4), 416–427, 1981. https://doi.org/10.1177/002224378101800402

4. Grewal D., Krishnan R., Baker J., & Borin N. The effect of store name, brand name and price discounts on consumers' evaluations and purchase intentions. Journal of Retailing, 74(3), 331–352, 1998. https://doi.org/10.1016/S0022-4359(99)80099-2

5. Guadagni P. M., & Little J. D. C. A logit model of brand choice calibrated on scanner data. Marketing Science, 2(3), 203–238, 1983. https://doi.org/10.1287/mksc.2.3.203

6. Jones R. M. The economic determinants of clothing consumption in the UK 1987–2000. Journal of Fashion Marketing and Management, 6(1), 8–20, 2002. https://doi.org/10.1108/13612020210448628

7. Kahneman D., & Tversky A. Prospect theory: An analysis of decision under risk. Econometrica, 47(2), 263–291, 1979. https://doi.org/10.2307/1914185

8. Parasuraman A., Zeithaml V. A., & Berry L. L. SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing, 64(1), 12–40, 1988.

9. Putler D. S. Incorporating reference price effects into a theory of consumer choice. Marketing Science, 11(3), 287–309, 1992. https://doi.org/10.1287/mksc.11.3.287

UDC 341.231.14:347.962

Ayana Azamatova

Master's Student of the Faculty of Law,
Ala-Too International University
(Bishkek, Kyrgyz Republic)

THE RIGHT TO A FAIR TRIAL IN THE KYRGYZ REPUBLIC: INTERNATIONAL STANDARDS, COMPARATIVE ASIAN PERSPECTIVES, AND STRATEGIC PATHWAYS TO REFORM

Abstract: The right to a fair trial is a fundamental human right and the cornerstone of a democratic society based on the rule of law. Despite the integration of international human rights standards into the national legislation of the Kyrgyz Republic, public trust in the judiciary remains critically low. This article examines the systemic challenges to achieving a fair trial in Kyrgyzstan by analyzing high-profile judicial cases that highlighted significant procedural violations. To identify viable solutions, the study provides a comparative analysis of judicial reforms in other Asian jurisdictions, including Japan, South Korea, Singapore, and Kazakhstan. Furthermore, it explores specific, actionable methods essential for building an independent and transparent judicial system. These methods include the expansion of jury trials, the empowerment of the defense bar, the role of public scrutiny, and the comprehensive implementation of digital technologies, such as automated case distribution and audio-video recording (AVR). The research concludes that transitioning from declaratory legal norms to actual justice requires profound institutional autonomy, digital transparency, and active civil society oversight.

Keywords: fair trial, Kyrgyz Republic, UN standards, judicial reform, transparency, Asian legal systems, jury trial, audio-video recording, human rights, public trust.

Introduction. The right to a fair trial is universally recognized as the ultimate safeguard of human rights and the mechanism through which the rule of law is enforced. Enshrined in Article 10 of the Universal Declaration of Human Rights (UDHR) and Article 14 of the International Covenant on Civil and Political Rights (ICCPR), this right ensures that everyone is entitled to a fair and public hearing by a competent, independent, and impartial tribunal established by law [1, p. 45]. In the Kyrgyz Republic, the Constitution formally guarantees these principles, reflecting the nation's commitment to international human rights frameworks [2]. However, the practical realization of the right to a fair trial remains one of the most acute and enduring challenges in the country, characterized by a persistent deficit of public trust in the judiciary, allegations of political interference, and widespread procedural violations.

The United Nations Human Rights Committee has repeatedly emphasized that the right to a fair trial is not merely about the formal existence of procedural laws, but about the actual, functional independence of judges and the equality of arms between the prosecution and the defense [3, p. 112]. In Kyrgyzstan, while the legislative framework is largely aligned with UN standards, implementation frequently falters. Sociological studies and public opinion polls consistently demonstrate that a significant portion of the Kyrgyz population views the courts as corrupt, biased, or subservient to the executive branch and law enforcement agencies [4, p. 78].

The Crisis of Trust: High-Profile Cases in Kyrgyzstan. This crisis of confidence is deeply rooted in a history of high-profile cases where the principles of a fair trial were visibly compromised, drawing international condemnation. One of the most glaring examples at the international level is the case of human rights defender Azimjon Askarov. Following the tragic ethnic clashes in southern Kyrgyzstan in 2010, Askarov was sentenced to life imprisonment in a trial marred by credible allegations of torture, lack of access to legal counsel, and severe courtroom intimidation by hostile crowds. In 2016, the UN Human Rights Committee officially concluded that Kyrgyzstan had violated Askarov's rights under Article 14 of the ICCPR, explicitly stating that he was denied a fair trial [5, p. 14]. Despite the UN's binding request for his release and a retrial, the national courts upheld his sentence, and he eventually died in prison in 2020. This case served as a stark demonstration of the systemic barriers to justice when judicial mechanisms succumb to political and public pressure.

More recently, the "Kempir-Abad case" (2022–2024) highlighted the pervasive misuse of closed-door trials. Over twenty politicians, activists, and civil society leaders were arrested on charges of plotting mass riots after protesting a border agreement with Uzbekistan [6, p. 22]. The judicial proceedings were heavily criticized by international human rights watchdogs because key hearings were held behind closed doors. The authorities cited "state secrets" as the rationale. However, under international law, the exclusion of the public and the media from court proceedings must be strictly limited to exceptional circumstances. The systematic closure of the Kempir-Abad hearings violated the principle of an open court, depriving the accused of public scrutiny, which is a vital safeguard against arbitrary justice [7, p. 156].

Similarly, the persecution of investigative journalists, such as the deportation of Bolot Temirov and the prosecution of the "11 media workers" from *Temirov Live* in 2024, showcased how the judicial system can be weaponized to suppress free speech [8, p. 30]. In these cases, defense lawyers frequently reported that their petitions were summarily dismissed by judges, violating the principle of adversarial proceedings and the equality of arms.

Comparative Perspective: Judicial Models in Asia. To understand how Kyrgyzstan can develop its judicial system, it is highly instructive to examine how other Asian nations have tackled similar issues of public mistrust and judicial inefficiency. Asia presents a diverse spectrum of legal cultures, offering valuable lessons in reform.

Japan and South Korea: Both of these robust Asian democracies faced periods where the public felt disconnected from a highly bureaucratic and elite judiciary. To democratize justice and inject public common sense into the courts, both countries introduced citizen participation systems. In 2009, Japan introduced the *Saiban-in* (lay judge) system for serious criminal trials, where six randomly selected citizens sit alongside three professional judges to determine guilt and sentencing [9, p. 102]. South Korea implemented a similar advisory jury system in 2008. Studies show that these reforms drastically increased judicial transparency and public trust in both nations, as professional judges were forced to explain legal concepts in accessible language, and backdoor deals became nearly impossible [10, p. 55].

Singapore: Singapore offers a different model, focusing on extreme efficiency, high remuneration, and zero tolerance for corruption. The Singaporean judiciary is renowned globally for its technological advancement, utilizing e-litigation platforms that make case management transparent and fast [11, p. 210]. Furthermore, judges in Singapore are paid salaries comparable to top corporate lawyers, effectively eliminating financial incentives for bribery.

While criticized by some for its strict defamation laws, Singapore proves that a technologically advanced, well-funded judiciary can eradicate petty corruption and procedural delays.

Kazakhstan: A compelling example from a neighboring Central Asian state is Kazakhstan's approach to bypassing domestic judicial distrust in the corporate sector. Recognizing that foreign investors did not trust local courts, Kazakhstan established the Astana International Financial Centre (AIFC) Court [12, p. 88]. This court operates entirely on English common law, uses the English language, and employs independent British judges. While this is an extreme measure (essentially "outsourcing" justice), it highlights the profound economic cost of a mistrusted judiciary and the drastic steps states must take to guarantee a fair trial to attract investment.

Strategic Methods for Achieving a Fair Trial in Kyrgyzstan. Drawing from international standards, the negative lessons of domestic high-profile cases, and successful Asian reforms, Kyrgyzstan must adopt concrete, systemic methods to guarantee the right to a fair trial.

1. Expanding and Empowering Jury Trials (Суд присяжных) Borrowing from the experiences of Japan and South Korea, Kyrgyzstan must fully implement and expand the institution of the jury trial. Currently, the law provides for jury trials only in a highly restricted category of crimes (e.g., life imprisonment cases), and its implementation has been repeatedly delayed. A jury trial breaks the monolithic power of the state prosecutor and the judge. When ordinary citizens decide on guilt, it is much harder for political figures to dictate the outcome via a phone call (so-called "telephone justice"). Expanding jury trials to a broader range of serious crimes is the most direct method to democratize the courts and restore public faith [13, p. 144].

2. Radical Transparency and Public Resonance (Огласка) In transitional democracies where institutional independence is weak, public scrutiny acts as the primary external check on judicial arbitrariness. When journalists, human rights defenders, and ordinary citizens attend court hearings and report on them, it significantly increases the political and reputational cost of delivering an unjust verdict. The state must unconditionally guarantee the right of the public and the media to attend court sessions. The practice of arbitrarily declaring trials "closed" must be outlawed, except in explicitly defined cases of severe state security or the protection of minors. The Supreme Court should issue a binding plenary ruling clarifying that media presence is a right, not a privilege granted at the whim of the presiding judge.

3. Deep Digitalization: AVR and Automated Case Distribution The introduction of the Audio and Video Recording (AVR) system in Kyrgyz courts was a massive step intended to eliminate the falsification of written court protocols [14, p. 89]. However, lawyers frequently complain that judges intentionally turn off the AVR systems during crucial moments, citing "technical malfunctions," or conduct important procedural steps in private chambers [15, p. 210]. To achieve a fair trial, the procedural codes must be amended to explicitly state that any hearing conducted without AVR—unless there is a verified, catastrophic power grid failure—is inherently invalid, and resulting decisions must be annulled.

Furthermore, digitalization must extend to the automated distribution of cases. Currently, court chairpersons often assign sensitive cases to "loyal" judges. By implementing a blockchain-based or AI-driven automated case assignment system, the system can eliminate "forum shopping" and ensure that case distribution is entirely blind and impartial [16, p. 77].

4. Institutional Parity: Strengthening the Bar (Адвокатура) A fair trial is impossible without the equality of arms. In Kyrgyzstan, the prosecutor represents the formidable apparatus of the state, while defense lawyers often face procedural hurdles, lack of access to evidence, and sometimes even physical intimidation. To combat this, the status of the legal profession (Advocacy) must be elevated. Defense attorneys must be granted equal rights in gathering evidence, and any obstruction of a lawyer's work by law enforcement should be strictly criminalized.

5. Financial and Institutional Autonomy Finally, as long as judges rely on the executive branch for their tenure, promotions, and budget, political loyalty will invariably trump the law in high-stakes cases. Following the Singaporean model, the judiciary must be granted full fiscal autonomy, managing its own budget without executive interference. Furthermore, the selection and dismissal of judges must be entirely in the hands of an independent Council of Judges, free from the influence of the Presidential Administration or Parliament.

Conclusion. The right to a fair trial in the Kyrgyz Republic cannot be achieved simply by rewriting laws or declaring adherence to UN conventions; it requires a radical shift in legal culture and state practice. High-profile cases over the last decade have repeatedly demonstrated that legal norms are easily bypassed when institutional independence and transparency are lacking. To build a system comparable to the leading jurisdictions in Asia, Kyrgyzstan must open its courtroom doors to the public. Civil society must be allowed to fulfill its watchdog role, jury trials must be expanded to empower citizens, and courtroom technologies like audio-video recording and automated case distribution must be strictly enforced. Only through maximum transparency, public accountability, and an unwavering commitment to equality of arms can Kyrgyzstan develop a judicial system that truly serves justice rather than power.

References:

1. Nowak, M. U.N. Covenant on Civil and Political Rights: CCPR Commentary. 2nd ed. / M. Nowak. – Kehl: N.P. Engel, 2005. – 1278 p.

2. Конституция Кыргызской Республики (принята референдумом 11 апреля 2021 года) // Эркин-Тоо. – 2021. – 5 мая (№ 39).

3. Трехсвятская, Е. М. Международные стандарты справедливого судебного разбирательства и их реализация в национальном праве / Е. М. Трехсвятская // Вестник международного права. – 2019. – № 4. – С.

4. Сыдыков, А. А. Проблемы доверия граждан к судебной системе Кыргызской Республики / А. А. Сыдыков // Право и государство: теория и практика. – 2021. – № 8(200). – С. 77-80.

5. United Nations Human Rights Committee. Views adopted by the Committee under article 5 (4) of the Optional Protocol, concerning communication No. 2231/2012 (Azimjan Askarov v. Kyrgyzstan). – Geneva: UN, 2016. – 18 p.

6. Ибраимова, Г. Анализ Кемпир-Абадского дела / Г. Ибраимова // Центральноазиатский юридический журнал. – 2024. – Т. 5, № 1.С. 20-35.

7. O'Flaherty, M. Human Rights and the UN: Practice Before the Treaty Bodies / M. O'Flaherty. – London: Sweet & Maxwell, 2002. – 256 p.

8. Freedom House. Freedom in the World 2024: The Mounting Damage of Flawed Elections and Armed Conflict – Kyrgyzstan Country Report. – Washington, D.C.: Freedom House, 2024. – 45 p.

9. Matsuo, K. The Saiban-in System: Japan's New Criminal Trial System / K. Matsuo // Asian Journal of Criminology. – 2010. – Vol. 5, No. 2. – P. 101-115.

10. Kim, J. Citizen Participation in Criminal Trials in South Korea / J. Kim // International Journal of Law, Crime and Justice. – 2018. – Vol. 52. – P. 50-61.

11. Menon, S. Technology and the Delivery of Justice in Singapore / S. Menon // Singapore Journal of Legal Studies. – 2020. – P. 205-225.

12. Suleimenov, M. K. The Astana International Financial Centre: A New Era for Dispute Resolution in Central Asia / M. K. Suleimenov // Asian Dispute Review. – 2019. – Vol. 21. – P. 85-91.

13. Осмонов, Д. К. Перспективы развития суда присяжных в уголовном судопроизводстве Кыргызстана / Д. К. Осмонов // Вестник КРСУ. – 2022. – Т. 22, № 5. – С. 142-147.

14. Джумабекова, А. К. Цифровизация правосудия в Кыргызской Республике: проблемы аудио- и видеофиксации судебных процессов / А. К. Джумабекова // Вестник КНУ им. Ж. Баласагына. – 2023. – № 2. – С. 88-94.

15. Сманалиев, К. М. Уголовный процесс Кыргызской Республики: учебник / К. М. Сманалиев. – Бишкек: Норма, 2020. – 432 с.

16. Aliev, R. Artificial Intelligence in Court Administration: Experiences from Central Asia / R. Aliev // Journal of Digital Law and Policy. – 2023. – Vol. 4. – P. 72-85.

УДК 373:37.09

Нургалиев Айдос Ерланулы

DBA докторант
Farabi International Business School
Казахский Национальный Университет имени аль-Фараби
(г.Алматы, Казахстан)

Абинова Альфия Юрьевна

PhD, ассоц.проф.
Farabi International Business School
Казахский Национальный Университет имени аль-Фараби
(г.Алматы, Казахстан)

UNIT ECONOMICS ЧАСТНОЙ ШКОЛЫ В КАЗАХСТАНЕ: СЕБЕСТОИМОСТЬ ОБУЧЕНИЯ, МАРЖИНАЛЬНОСТЬ И ФАКТОРЫ ФИНАНСОВОЙ УСТОЙЧИВОСТИ

Аннотация. В статье рассматривается финансово-управленческая модель частной средней школы в Казахстане через призму unit economics. Актуальность исследования обусловлена ростом частного сектора среднего образования, увеличением численности школьников в крупных городах, развитием государственного образовательного заказа и переходом к цифровым механизмам финансирования, включая ваучерную модель. Частная школа в современных условиях выступает не только образовательной организацией, но и сложной экономической системой, где качество обучения напрямую связано с управлением себестоимостью, загрузкой классов, удержанием контингента и структурой доходов.

Цель статьи - разработать прикладную модель расчёта себестоимости обучения одного ученика, определить ключевые драйверы маржинальности и предложить систему KPI финансовой устойчивости частной школы. Методологической основой выступают activity-based costing, расчёт cost per student, анализ постоянных и переменных затрат, сценарное моделирование загрузки, а также показатели unit economics: CAC, LTV, retention и occupancy rate.

Для защиты коммерческой информации в статье используется нормализованный модельный кейс частной школы полного дня в крупном городе Казахстана с проектной численностью 500 учеников. Представленные цифры не являются финансовой отчётностью конкретной организации, а отражают типовую управленческую конфигурацию частной школы. В базовой модели доходов государственный образовательный заказ не включён в основную выручку, поскольку участие частной школы в ГОЗ зависит от региона, контингента, подтверждённых данных, условий финансирования и статуса образовательной организации. Вместе с тем в статье отдельно рассчитан возможный эффект ГОЗ как сценарный фактор финансовой устойчивости.

Согласно базовому модельному расчёту, при совокупных годовых доходах 1 250 млн тенге и расходах 1 080 млн тенге операционная прибыль составляет 170 млн тенге, а

операционная маржинальность - 13,6%. Себестоимость одного ученика составляет 2,16 млн тенге в год. При полном участии в ГОЗ по нормативам 2026 года для модельной школы на 500 учеников потенциальный объём государственного финансирования может составить около 278,2 млн тенге в год.

Ключевые слова: частная школа, unit economics, себестоимость обучения, cost per student, маржинальность, retention, LTV, CAC, государственный образовательный заказ, ваучерное финансирование, Казахстан.

Введение. В последние годы частный сектор среднего образования в Казахстане становится важным элементом образовательной инфраструктуры. Рост численности школьников в крупных городах и регионах создаёт дополнительную нагрузку на государственные школы и одновременно формирует спрос на частные образовательные организации. По данным OrtaBilim, на 19.05.2026 в Казахстане отражены 9 237 организаций образования, из них 1 438 - частные организации образования. По городу Алматы указано 541 организация образования, включая 259 частных организаций образования.

Дополнительно динамику спроса подтверждают данные по численности школьников в крупных городах. В 2024-25 учебном году в Алматы числились 337 083 школьника, в Астане - 286 633, в Шымкенте -249 748. Эти показатели показывают, что частные школы развиваются не только как альтернатива государственным школам, но и как один из инструментов снижения инфраструктурной нагрузки на систему среднего образования.

Институциональная среда также меняется. В материалах по ваучеризации государственного образовательного заказа указано, что в 2026 году сохраняется финансирование через сметы и ГОЗ, а с 2027 года планируется переход к финансированию по ваучерам. В презентационных материалах также указано, что всего в Казахстане насчитывается 7 732 школы, из них 6 887 - государственные и 845 - частные. Это означает, что частные школы постепенно включаются в более широкую систему государственного финансирования, цифрового мониторинга и прозрачной отчётности.

В этих условиях частная школа работает одновременно как образовательная организация и как экономическая система. С одной стороны, родители ожидают высокого качества обучения, сильных педагогов, безопасной среды, полноценного питания, дополнительных кружков, развития английского языка, математики, финансовой грамотности и индивидуального сопровождения ребёнка. С другой стороны, школа должна сохранять финансовую устойчивость: обеспечивать своевременную оплату труда, покрывать расходы на здание, коммунальные услуги, налоги, безопасность, питание, маркетинг, повышение квалификации педагогов и развитие инфраструктуры.

Проблема заключается в том, что многие частные школы принимают управленческие решения на основе общей выручки и совокупных расходов, не рассчитывая детально себестоимость одного ученика, точку безубыточности, влияние недобора, стоимость привлечения нового ученика и экономический эффект удержания семей. В результате школа может внешне демонстрировать рост контингента, но фактически терять финансовую устойчивость из-за высокой доли фиксированных расходов, роста фонда оплаты труда, затрат на питание, аренду, безопасность и маркетинг.

Особое значение приобретает вопрос государственного образовательного заказа. ГОЗ может стать важным источником финансовой устойчивости частной школы, однако его нельзя механически включать в базовую структуру доходов без уточнения условий. Объём ГОЗ зависит от количества учеников, уровня образования, региона, нормативов подушевого финансирования, наличия доплат, статуса школы и подтверждённого контингента. Поэтому в настоящей статье ГОЗ рассматривается не как гарантированная строка выручки, а как отдельный сценарный фактор.

Именно поэтому для частных школ становится важным переход к модели unit economics. Такой подход позволяет рассматривать одного ученика или одну семью как базовую экономическую единицу анализа. В этом случае директор школы может видеть не только общую выручку, но и реальную себестоимость одного ученика, маржу, срок окупаемости маркетинговых расходов, влияние retention на LTV и критическую загрузку классов.

Цель статьи: разработать и обосновать модель unit economics частной школы в Казахстане на основе расчёта себестоимости обучения, маржинальности и факторов финансовой устойчивости.

Для достижения цели поставлены следующие задачи:

1. Описать рыночный и институциональный контекст развития частных школ в Казахстане.
2. Предложить структуру расчёта себестоимости обучения одного ученика.
3. Выделить основные драйверы затрат и маржинальности частной школы.
4. Рассчитать возможный сценарный эффект ГОЗ на финансовую устойчивость.
5. Провести сценарный анализ влияния загрузки школы на маржинальность.
6. Сформировать систему KPI для управленческого мониторинга частной школы.

Методы и материалы. Исследование выполнено в формате прикладного управленческо-экономического анализа с элементами case study и сценарного моделирования. В качестве объекта анализа рассматривается условная частная средняя школа полного дня, осуществляющая образовательную деятельность с 1 по 11 класс в крупном городе Казахстана.

Для сохранения конфиденциальности коммерческих данных в статье используется нормализованный модельный кейс. Абсолютные финансовые показатели округлены и агрегированы. Они не являются бухгалтерской или управленческой отчётностью конкретной школы, а используются для демонстрации методики расчёта unit economics. При этом сохранена типовая логика структуры доходов и расходов частной школы полного дня: фонд оплаты труда выступает крупнейшим драйвером себестоимости, питание и развозка имеют преимущественно сервисную природу, а дополнительное образование может быть одним из источников операционной маржинальности.

Методика исследования объединяет четыре подхода:

1. Activity-Based Costing (ABC-costing) - распределение затрат по видам деятельности.
2. Cost per student analysis - расчёт себестоимости обучения одного ученика.
3. Scenario analysis - моделирование влияния загрузки школы и ГОЗ на маржинальность.
4. Unit economics framework - анализ CAC, LTV, retention, occupancy rate и операционной маржи.

ABC-costing позволяет распределять расходы школы не только по бухгалтерским статьям, но и по видам деятельности, которые создают образовательную услугу. Для частной школы такой подход особенно важен, поскольку стоимость обучения включает не только проведение уроков, но и организацию полного дня, питание, безопасность, воспитательную работу, кружки, методическую поддержку, коммуникацию с родителями и административное сопровождение.

В рамках исследования затраты частной школы распределены по следующим блокам:

- основное обучение;
- администрирование;
- инфраструктура;
- безопасность;
- питание;
- развозка;
- дополнительное образование;
- маркетинг и продажи.

Для расчёта unit economics используются следующие показатели:

Себестоимость одного ученика:

Cost per student = Total annual operating costs / Average annual number of students

Доход на одного ученика:

Revenue per student = Total annual revenue / Average annual number of students

Операционная маржа:

Operating margin = Operating profit / Total revenue

Стоимость привлечения ученика:

CAC = Total marketing and sales expenses / Number of newly enrolled students

Долгосрочная ценность семьи:

LTV = Annual revenue per family × Average retention period × Operating margin

Загрузка проектной мощности:

Occupancy rate = Actual number of students / Project capacity

Потенциальный объём ГОЗ:

State order funding = Σ(Number of students by level × Per-student funding norm by level) + utility funding norm × total number of students

В расчёте ГОЗ использованы нормативы подушевого финансирования на 2026 год для города без экологических и радиационных доплат:

- 1-4 классы - 416 092 тенге на одного ученика в год;
- 5-9 классы - 575 783 тенге на одного ученика в год;
- 10-11 классы - 691 840 тенге на одного ученика в год;
- коммунальные затраты для второй группы регионов, включая города Астана, Алматы и Шымкент, - 19 463 тенге на одного ученика.

Для модельной школы принято следующее распределение контингента:

- 1-4 классы - 180 учеников;
- 5-9 классы - 240 учеников;
- 10-11 классы - 80 учеников;
- всего - 500 учеников.

Результаты

1)Частная школа как экономическая система. Анализ показывает, что частная школа является организацией с высокой долей постоянных и условно-постоянных затрат. В отличие от многих видов бизнеса, школа не может гибко сокращать расходы пропорционально снижению количества учеников. Даже при недоборе контингента сохраняются расходы на здание, охрану, администрацию, коммунальные услуги, минимальный педагогический состав, лицензирование, санитарную безопасность и базовую инфраструктуру.

Следовательно, ключевым фактором финансовой устойчивости частной школы является не только размер родительской оплаты, но и уровень загрузки проектной мощности. При снижении количества учеников себестоимость одного ученика увеличивается, поскольку постоянные расходы распределяются на меньшее число обучающихся.

2)Базовая нормализованная модель доходов без учёта ГОЗ. Для демонстрации методики была построена модель частной школы полного дня на 500 учеников. В базовой модели ГОЗ не включён в основную выручку, чтобы не смешивать коммерческую модель школы и сценарное государственное финансирование.

Таблица 1. Нормализованная структура доходов модельной частной школы без учёта ГОЗ.

Источник дохода	**Сумма, млн тг**	**Доля**
Родительская оплата за обучение	925	74%
Питание	100	8%
Развозка	75	6%
Дополнительные образовательные услуги	125	10%
Прочие доходы	25	2%
Итого доходы	1 250	100%

Основной источник дохода частной школы - родительская оплата за обучение. Однако для финансовой устойчивости важна диверсификация доходов. Питание и развозка чаще имеют сервисную природу и не всегда являются высокомаржинальными направлениями. Дополнительное образование, напротив, может создавать заметный вклад в операционную прибыль, если оно грамотно встроено в образовательную модель школы.

Важно подчеркнуть, что исключение ГОЗ из базовой таблицы не означает его незначительности. Напротив, ГОЗ может существенно влиять на экономику частной школы, но его необходимо рассчитывать отдельно, по нормативам и фактическому контингенту.

3) Нормализованная структура расходов.

Таблица 2. Нормализованная структура расходов модельной частной школы.

Статья расходов	**Сумма, млн тг**	**Доля от расходов**
ФОТ, налоги и социальные отчисления	560	52%
Питание	90	8%
Развозка	75	7%
Коммунальные услуги и содержание здания	85	8%
Учебные материалы, инвентарь, хозяйственные расходы	55	5%
Безопасность, IT, связь, лицензирование, аккредитация	50	5%

Маркетинг и продажи	35	3%
Амортизация / аренда / финансовая нагрузка	100	9%
Прочие операционные расходы	30	3%
Итого расходы	1 080	100%

Наиболее крупной статьёй расходов является фонд оплаты труда с налогами и социальными отчислениями. Это закономерно для образовательной организации, где качество услуги напрямую зависит от профессионального уровня педагогов, стабильности кадров и эффективности управленческой команды.

Вторым значимым блоком являются инфраструктурные расходы: коммунальные услуги, содержание здания, амортизация, аренда или финансовая нагрузка. Эти расходы имеют преимущественно постоянный характер и особенно сильно влияют на себестоимость при недозагрузке школы.

Питание и развозка также занимают заметную долю расходов. Они повышают ценность школы для родителей, но требуют точного управленческого расчёта. Если школа не контролирует закупки, логистику и операционные нормы, эти сервисы могут снижать общую маржинальность.

4) Расчёт cost per student и маржинальности. На основе нормализованной модели совокупные доходы школы составляют 1 250 млн тенге в год, совокупные расходы - 1 080 млн тенге, операционная прибыль - 170 млн тенге.

Таблица 3. Основные показатели unit economics без учёта ГОЗ.

Показатель	**Значение**
Количество учеников	500
Совокупные доходы	1 250 млн тг
Совокупные расходы	1 080 млн тг
Операционная прибыль	170 млн тг
Операционная маржинальность	13,6%
Revenue per student	2 500 000 тг/год
Cost per student	2 160 000 тг/год
Profit per student	340 000 тг/год

Расчёты показывают, что даже при годовой выручке 1,25 млрд тенге операционная маржинальность составляет 13,6%. Это демонстрирует, что частная школа не является бизнесом с автоматически высокой прибыльностью. Значительная часть доходов направляется на обеспечение качества, безопасности, кадровой устойчивости и инфраструктуры.

Себестоимость одного ученика составляет 2,16 млн тенге в год, или примерно 180 тыс. тенге в месяц при распределении на 12 месяцев. Данный показатель объясняет, почему частная школа полного дня не может устойчиво работать при слишком низкой стоимости обучения, если она обеспечивает полноценное питание, сильный кадровый состав, безопасность и дополнительные образовательные возможности.

5) Сценарный расчёт ГОЗ по нормативам 2026 года. Отдельно был рассчитан потенциальный объём государственного образовательного заказа для модельной школы на 500 учеников. Расчёт выполнен на основе нормативов подушевого финансирования на 2026 год для города без экологических и радиационных доплат.

Таблица 4. Расчёт потенциального ГОЗ для модельной школы на 500 учеников.

Уровень образования	Количество учеников	Норматив на 1 ученика в год, тг	Сумма, тг
1–4 классы	180	416 092	74 896 560
5–9 классы	240	575 783	138 187 920
10–11 классы	80	691 840	55 347 200
Итого операционное подушевое финансирование	500		268 431 680
Коммунальные затраты	500	19 463	9 731 500
Итого потенциальный объём ГОЗ в год			278 163 180

Расчёт показывает, что при полном участии модельной школы в ГОЗ потенциальный объём государственного финансирования может составить около 278,2 млн тенге в год. Это значительно больше условных 25 млн тенге, которые могут быть характерны только для частичного финансирования ограниченного количества учеников.

Следовательно, в структуре доходов частной школы ГОЗ не должен автоматически отражаться как 2% от выручки, если речь идёт о финансировании всего контингента. Более корректный подход - рассматривать ГОЗ через отдельные сценарии:

Таблица 5. Сценарии участия частной школы в ГОЗ.

Сценарий	Описание	Ориентировочный объём ГОЗ
Без ГОЗ	Школа работает только на родительской оплате и коммерческих доходах	0 тг
Частичный ГОЗ	Финансирование ограниченного количества учеников, например 40–50 человек	около 22–28 млн тг
Полный ГОЗ	Финансирование всего модельного контингента 500 учеников	около 278,2 млн тг

ГОЗ может использоваться по-разному в зависимости от модели школы: как фактор снижения родительской нагрузки, как источник компенсации части себестоимости или как механизм повышения финансовой предсказуемости. Однако в любом случае он требует прозрачного учёта, подтверждения контингента, соответствия требованиям и готовности школы к мониторингу.

6) Сценарии загрузки и влияние на маржинальность. Одним из ключевых факторов устойчивости является загрузка проектной мощности. Для частной школы с высокой долей постоянных затрат недобор учеников резко увеличивает себестоимость одного ученика.

Таблица 6. Сценарии загрузки школы.

Сценарий	Количество учеников	Уровень загрузки	Управленческий эффект
Пессимистичный	400	80%	Рост себестоимости на ученика, давление на маржу

Базовый	500	100%	Операционная устойчивость при контроле расходов
Оптимистичный	550	110%	Рост маржинальности при условии сохранения качества

Пессимистичный сценарий показывает, что снижение контингента до 400 учеников может существенно ухудшить экономику школы. Постоянные расходы сохраняются, а доходы снижаются. В такой ситуации школа может столкнуться с необходимостью повышать стоимость обучения, сокращать расходы или активнее развивать дополнительные источники дохода.

Базовый сценарий на 500 учеников обеспечивает устойчивую операционную модель. При такой загрузке школа способна покрывать основные расходы, инвестировать в качество и сохранять разумную маржинальность.

Оптимистичный сценарий на 550 учеников может повысить прибыльность, но несёт риск снижения качества при неправильном управлении. Если рост контингента приводит к переполнению классов, перегрузке педагогов, ухудшению сервиса и росту недовольства родителей, краткосрочный финансовый выигрыш может привести к долгосрочному снижению retention.

Обсуждение. Результаты исследования показывают, что высокая стоимость обучения не означает автоматически высокой прибыльности. Родительская оплата покрывает не только уроки, но и широкий комплекс обязательных процессов: оплату труда педагогов, налоги, питание, охрану, коммунальные услуги, санитарную безопасность, IT-инфраструктуру, методическое развитие, административное сопровождение и маркетинг.

На основе анализа можно выделить пять основных драйверов себестоимости частной школы.

Первый драйвер - фонд оплаты труда. Качество образования невозможно обеспечить без квалифицированных педагогов и стабильной управленческой команды. Поэтому ФОТ не должен рассматриваться только как статья расходов. Это инвестиция в качество, репутацию и удержание семей.

Второй драйвер - загрузка проектной мощности. Чем ниже загрузка, тем выше себестоимость одного ученика. Это делает retention одним из важнейших финансовых показателей школы.

Третий драйвер - инфраструктура. Здание, коммунальные расходы, безопасность, ремонт и амортизация формируют значительную постоянную нагрузку.

Четвёртый драйвер - сервис полного дня. Питание, развозка, кружки и сопровождение повышают ценность школы для родителей, но требуют точного расчёта.

Пятый драйвер - маркетинг и привлечение учеников. Если школа теряет учеников, ей приходится тратить больше средств на привлечение новых. Поэтому удержание семей экономически выгоднее постоянной компенсации оттока через рекламу.

Отдельного внимания заслуживает ГОЗ. Исправленный расчёт показывает, что государственное финансирование может быть не символической, а значимой частью экономики частной школы. При полном охвате модельного контингента 500 учеников объём ГОЗ может составить около 278,2 млн тенге в год. Это означает, что участие в ГОЗ

способно заметно повлиять на финансовую устойчивость, снизить давление на родительскую оплату или усилить возможности школы по развитию качества.

Однако ГОЗ не должен рассматриваться как автоматическая прибыль. Это целевое финансирование, связанное с образовательной услугой, подтверждением контингента, требованиями к прозрачности, цифровому учёту, соответствию лицензии, безопасности и кадровой укомплектованности. Поэтому в unit economics частной школы ГОЗ должен учитываться отдельно: не как универсальная доля выручки, а как сценарный и институционально обусловленный фактор.

Развитие государственного образовательного заказа и ваучерного финансирования меняет экономику частной школы. В новой модели финансовая устойчивость будет зависеть не только от родительской оплаты, но и от способности школы работать в цифровой системе: подтверждать фактический контингент, вести корректный учёт, взаимодействовать с LMS, проходить мониторинг и соответствовать требованиям.

Для частных школ это создаёт как возможности, так и риски. Возможность заключается в повышении предсказуемости финансирования и расширении доступности частного образования. Риск связан с ростом требований к прозрачности, отчётности и управленческой дисциплине. Следовательно, unit economics частной школы должен учитывать не только внутреннюю финансовую модель, но и институциональные механизмы финансирования.

Оптимизация себестоимости частной школы не должна сводиться к простому сокращению расходов. В образовании чрезмерная экономия может привести к снижению качества, уходу сильных педагогов, ухудшению сервиса и росту оттока родителей. Более корректная управленческая логика заключается в управлении балансом между себестоимостью, ценой, государственным финансированием и воспринимаемой ценностью для родителей.

Для регулярного управленческого мониторинга частной школе рекомендуется использовать набор KPI, который отражает не только бухгалтерский результат, но и реальные факторы устойчивости.

Таблица 7. KPI финансовой устойчивости частной школы.

KPI	Формула	Управленческий смысл
Cost per student	Общие расходы / средний контингент	Себестоимость одного ученика
Revenue per student	Общие доходы / средний контингент	Доход на одного ученика
Operating margin	Операционная прибыль / доходы	Маржинальность проекта
Occupancy rate	Факт учеников / проектная мощность	Загрузка школы
Retention	Сохранившиеся ученики / база прошлого года	Удержание семей
CAC	Маркетинг и продажи / новые ученики	Стоимость привлечения ученика
LTV	Доход семьи × срок обучения × маржа	Долгосрочная ценность семьи
Payroll ratio	ФОТ / выручка	Кадровая нагрузка
Break-even enrollment	Постоянные расходы / маржа на ученика	Минимальный нужный контингент

Referral share	Заявки по рекомендациям / все заявки	Уровень доверия к школе
State funding per student	ГОЗ / количество профинансированных учеников	Размер государственного финансирования на ученика
ГОЗ coverage rate	Количество учеников по ГОЗ / общий контингент	Доля охвата государственным заказом

Эти показатели следует отслеживать ежемесячно или ежеквартально. Особенно важно связывать финансовые KPI с образовательными показателями: удовлетворённостью родителей, качеством обучения, дисциплиной, результатами экзаменов, олимпиадами и удержанием учеников.

В качестве практического инструмента предлагается использовать калькулятор unit economics. Он должен включать входные параметры: проектную мощность, фактический контингент, среднюю родительскую оплату, доходы от питания, развозки и дополнительного образования, ФОТ, налоги, аренду или амортизацию, коммунальные расходы, питание, охрану, маркетинг, IT, лицензирование, аккредитацию и прочие расходы.

Отдельный блок калькулятора должен быть посвящён ГОЗ. В него необходимо включить:

- количество учеников 1–4 классов;
- количество учеников 5–9 классов;
-количество учеников 10–11 классов;
-норматив на одного ученика по каждому уровню;
-коммунальный норматив;
-наличие или отсутствие дополнительных коэффициентов;
-общий объём потенциального государственного финансирования;
-долю ГОЗ в общей структуре доходов;
-влияние ГОЗ на родительскую оплату и маржинальность.

Выходными показателями калькулятора должны быть себестоимость одного ученика, доход на одного ученика, операционная прибыль, маржинальность, точка безубыточности, CAC, LTV, запас финансовой устойчивости, влияние недобора учеников на себестоимость, влияние retention на прибыль и влияние ГОЗ на общую финансовую модель.

Такой калькулятор позволяет директору школы моделировать последствия управленческих решений до их принятия. Например, можно заранее увидеть, как изменится маржинальность при снижении контингента на 10%, росте ФОТ на 15%, увеличении стоимости питания, запуске нового кружкового направления, изменении маркетингового бюджета или подключении части учеников к ГОЗ.

Заключение. Исследование показывает, что финансовая устойчивость частной школы определяется не только уровнем родительской оплаты, но и качеством управления себестоимостью, загрузкой проектной мощности, удержанием контингента, структурой дополнительных услуг, возможностью участия в государственном образовательном заказе и способностью школы адаптироваться к новым механизмам финансирования.

На основе нормализованной модели частной школы полного дня на 500 учеников установлено, что при совокупных годовых доходах 1 250 млн тенге и расходах 1 080 млн тенге операционная прибыль составляет 170 млн тенге, а операционная маржинальность

- 13,6%. Себестоимость одного ученика составляет 2,16 млн тенге в год, а доход на одного ученика - 2,5 млн тенге в год.

Исправленный сценарный расчёт ГОЗ показывает, что при полном охвате модельного контингента потенциальный объём государственного финансирования может составить около 278,2 млн тенге в год. Следовательно, ГОЗ не должен условно отражаться как незначительная доля выручки, если речь идёт о финансировании всего контингента. Более корректно рассматривать его как отдельный сценарный фактор, который может существенно менять экономику частной школы.

Данные результаты показывают, что даже при высокой совокупной выручке частная школа может иметь умеренную маржинальность из-за значительной доли обязательных расходов. Ключевыми факторами финансовой устойчивости являются фонд оплаты труда, загрузка классов, инфраструктурные расходы, питание, безопасность, дополнительные образовательные услуги, retention и институциональная готовность к работе с ГОЗ.

Практическая значимость статьи заключается в предложении KPI и unit-economics калькулятора, который может использоваться директорами частных школ для планирования бюджета, расчёта точки безубыточности, оценки сценариев загрузки, управления маркетингом, расчёта ГОЗ и принятия решений о развитии дополнительных услуг.

Главный вывод исследования состоит в следующем: частная школа финансово устойчива не тогда, когда имеет высокую стоимость обучения, а тогда, когда умеет точно управлять себестоимостью, ценностью, загрузкой, удержанием, качеством и возможностями государственного финансирования.

Список использованной литературы:

1. Министерство просвещения Республики Казахстан. Данные по частным организациям образования Республики Казахстан. Ответ на запрос №ЗТ-2026-00551754 от 27 февраля 2026 года.

2. АО «Информационно-учётный центр» Министерства финансов Республики Казахстан. Ваучеризация государственного образовательного заказа на среднее образование. Презентационные материалы OrtaBilim, 2026.

3. OrtaBilim / e-Qazyna. Открытые данные по организациям образования, школам и финансированию государственного образовательного заказа. Министерство финансов Республики Казахстан.

4. Расчёт объёма подушевого финансирования для частных школ на 2026 год. Нормативы подушевого финансирования по уровням образования и коммунальным затратам.

5. Cooper, R., & Kaplan, R. S. (1988). Measure Costs Right: Make the Right Decisions. Harvard Business Review.

6. Kaplan, R. S., & Anderson, S. R. (2007). Time-Driven Activity-Based Costing: A Simpler and More Powerful Path to Higher Profits. Harvard Business School Press.

7. Drury, C. (2018). Management and Cost Accounting. Cengage Learning.

8. Gupta, S., & Lehmann, D. R. (2005). Managing Customers as Investments: The Strategic Value of Customers in the Long Run. Wharton School Publishing.

9. Reichheld, F. F. (1996). The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value. Harvard Business School Press.
10. Zeithaml, V. A. (1988). Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence. Journal of Marketing, 52(3), 2–22.
11. Kotler, P., & Keller, K. L. (2016). Marketing Management. Pearson.
12. OECD. Education at a Glance. OECD Publishing.
13. UNESCO. Global Education Monitoring Report. UNESCO Publishing.
14. World Bank. Education Finance and Public-Private Partnerships in Education. World Bank Publications.

УДК 005.21:336.773

Глазинский Евгений Юрьевич

магистрант программы EMBA
Научный руководитель: Умирзаков Самажан Ынтыкбаевич
д.э.н., профессор
Университет Нархоз
(г. Алматы, Казахстан)

АГЕНТСКАЯ СЕТЬ И ДОВЕРИТЕЛЬНЫЕ КАНАЛЫ ПРОДАЖ КАК ИНСТРУМЕНТЫ РОСТА ССУДНОГО ПОРТФЕЛЯ МИКРОКРЕДИТНОЙ ОРГАНИЗАЦИИ (НА ПРИМЕРЕ АВТОЛОМБАРДА)

Аннотация: В статье на материалах кейса автоломбарда (ТОО «Ломбард Прогресс», г. Алматы) обосновывается роль агентской сети и доверительных каналов продаж как ключевых инструментов роста ссудного портфеля микрокредитной организации в условиях насыщенного и маржинально сжатого рынка. Проанализирована структура каналов привлечения клиентов, в которой агентский канал и повторные обращения формируют около 80 % сделок. Рассмотрена юнит-экономика агентского канала (вознаграждение 1–2 % при существенно более низкой стоимости привлечения по сравнению с платной рекламой) и его влияние на качество портфеля. Показана роль CRM-системы (Bitrix24) как инфраструктуры масштабирования каналов. Предложен комплекс инструментов наращивания сети и удержания клиентов, увязанный с целевыми показателями роста портфеля до 1,5 млрд тг при ROE ≥ 30 % и NPL ≤ 3 %.

Ключевые слова: микрокредитная организация, автоломбард, ссудный портфель, агентская сеть, доверительные каналы, лояльность клиентов, CRM, стратегия роста.

Введение

Ломбардный сегмент микрофинансового рынка Республики Казахстан вступил в фазу зрелости и обострения конкуренции. По данным Национального Банка РК, на 1 апреля 2026 года в стране действовало 466 ломбардов с совокупными активами 653,9 млрд тг и кредитным портфелем 481,3 млрд тг [1]. За первый квартал 2026 года активы сектора выросли почти на 12 %, а доля займов с просрочкой свыше 90 дней увеличилась с 1,6 до 2,1 % портфеля, что свидетельствует об усилении конкуренции и накоплении кредитного риска. Рынок неоднороден и делится на три сегмента: ломбарды, работающие с золотом и ювелирными изделиями (крупнейшие игроки — М-Ломбард, МК-Ломбард, МК-Золото Ломбард с активами 69–106 млрд тг), автоломбарды и организации смешанного типа (например, Сейф Ломбард — 52,0 млрд тг). При этом дифференциация по цене продукта внутри сегментов ограничена.

Сегмент автоломбардов — кредитования под залог автотранспорта — отличается особой плотностью конкуренции. ТОО «Ломбард Прогресс» с активами около 1,17 млрд тг и кредитным портфелем 1,01 млрд тг занимает 43-е место среди 466 ломбардов РК и входит в группу ключевых «чистых» автоломбардов г. Алматы [1]. Качество портфеля у игроков сегмента различается кардинально: доля проблемных займов (NPL 90+) у отдельных автоломбардов превышает 30 %, тогда как у лидеров по качеству не достигает

и 1 %. Ломбарды при этом работают исключительно с физическими лицами и не обязаны рассчитывать коэффициент долговой нагрузки заёмщика, что обеспечивает скорость выдачи как ключевое конкурентное преимущество перед МФО [2]. В этих условиях устойчивый рост ссудного портфеля определяется не ценой займа, а эффективностью каналов привлечения и удержания клиентов и качеством риск-отбора.

Цель настоящей статьи — на основе кейса ТОО «Ломбард Прогресс» (г. Алматы, автоломбард, работающий на рынке с 2011 года) обосновать, что в условиях насыщенного рынка ключевыми инструментами роста ссудного портфеля выступают агентская сеть и доверительные каналы продаж (повторные обращения и рекомендации), а CRM-система является необходимой инфраструктурой их масштабирования. Научная новизна состоит в систематизации каналов привлечения автоломбарда и оценке их вклада в рост портфеля с увязкой к целевым финансовым показателям.

Материалы и методы

Исследование выполнено в методологии единичного кейс-стади (case study). Эмпирическую базу составили официальные данные Национального Банка РК по ломбардному сектору (сведения о ломбардах по состоянию на 1 января и 1 апреля 2026 года), аудированная финансовая отчётность ТОО «Ломбард Прогресс» за 2025 год и внутренние управленческие данные компании: структура каналов привлечения клиентов, параметры агентского вознаграждения, доля повторных обращений, показатели доходности и качества портфеля. Применены методы качественного анализа бизнес-процессов и расчёта юнит-экономики каналов привлечения. Теоретической рамкой послужили концепции стратегий проникновения на рынок [3], управления маркетингом и каналами распределения [4], а также экономики клиентской лояльности [5].

Результаты и обсуждение

Структура каналов привлечения. Анализ сделок ТОО «Ломбард Прогресс» показывает, что агентский канал обеспечивает 40–50 % выдач, а повторные клиенты — ещё 30–40 %. Суммарно около 80 % сделок проходит через доверительные каналы — то есть через посредников, лично заинтересованных в качестве клиента, либо через клиентов, уже имеющих положительный опыт сотрудничества. Доля «холодного» привлечения (платная реклама, случайный трафик) остаётся второстепенной. Эта структура принципиально отличает бизнес-модель от моделей, опирающихся на массовый маркетинг, и определяет логику дальнейшего роста.

Юнит-экономика агентского канала. Агентская сеть организации насчитывает около десяти действующих агентов; вознаграждение составляет 1–2 % от суммы сделки. Ключевое преимущество канала — не только низкая удельная стоимость привлечения по сравнению с платной рекламой, но и предварительная «фильтрация» клиента: агент приводит платёжеспособного заёмщика, в котором лично заинтересован, что снижает вероятность дефолта и поддерживает качество портфеля. Это особенно значимо для сегмента автоломбардов, где, по данным Национального Банка РК, разброс доли проблемных займов между игроками превышает 30 процентных пунктов. Таким образом, агентский канал одновременно работает на два целевых показателя — рост портфеля и удержание NPL на низком уровне.

Доверительные каналы и лояльность. Повторные обращения (30–40 % сделок) представляют собой наиболее рентабельный канал: стоимость их привлечения близка к нулю, конверсия выше, а кредитная история клиента уже известна. Согласно концепции

лояльности, удержание существующего клиента кратно дешевле привлечения нового, а накопленная история взаимодействия повышает точность оценки риска [5]. Для автоломбарда это означает, что инвестиции в качество сервиса и клиентский опыт (прозрачные условия, отсутствие навязанных страховок как элемент уникального торгового предложения) напрямую конвертируются в рост портфеля за счёт повторных выдач и рекомендаций.

CRM как инфраструктура масштабирования. Развёрнутая в организации CRM-система Bitrix24 выполняет роль инфраструктуры, без которой доверительные каналы не масштабируются. Она обеспечивает учёт и атрибуцию сделок по агентам (корректный расчёт вознаграждения и оценку эффективности каждого агента), сегментацию клиентской базы и автоматические триггеры повторных обращений (напоминания, предложения по истечении срока займа), а также накопление данных для скоринга. По мере роста числа агентов и клиентов ручное администрирование становится узким местом, и именно цифровизация процессов позволяет расширять сеть без пропорционального роста административных расходов.

Финансовая рамка и целевые показатели. По аудированной финансовой отчётности за 2025 год (ТОО «Алмир Консалтинг») процентные доходы компании составили 382,3 млн тг, чистый процентный доход — 163,6 млн тг, чистая прибыль — 17,6 млн тг. Сформированного процентного дохода достаточно для покрытия операционных расходов и агентского вознаграждения (1–2 % от суммы сделки). По данным Национального Банка РК, на конец 2025 года ссудный портфель компании достиг 1 048 млн тг при доле проблемных займов (NPL 90+) 2,9 % и собственном капитале 173,9 млн тг, что соответствует рентабельности собственного капитала (ROE) около 10 % [1]. По состоянию на 1 апреля 2026 года портфель составил 1 010 млн тг, что соответствует 43-му месту среди 466 ломбардов РК. Стратегические ориентиры на 2026–2027 годы и роль рассмотренных каналов в их достижении представлены в таблице 1.

Таблица 1 – Целевые показатели роста и вклад каналов привлечения

Показатель	Факт 2025	Цель 2026–2027	Ведущий канал
Ссудный портфель, млн тг	1 048	1 500	Агенты, повторные
ROE, %	≈ 10	≥ 30	Низкий CAC каналов
NPL 90+, %	2,9	≤ 3	Фильтрация агентами
Агентская сеть, чел.	≈ 10	20–25	Развитие сети

Примечание – составлено авторами по данным ТОО «Ломбард Прогресс» и Национального Банка РК.

Предлагаемые инструменты роста. На основе проведённого анализа предлагается комплекс взаимоувязанных инструментов: (1) многоуровневая система мотивации агентов с прогрессивным вознаграждением за объём и качество приводимых сделок; (2) формализованная реферальная программа для действующих клиентов, переводящая «сарафанное радио» в управляемый канал; (3) автоматизация повторных продаж в CRM (триггерные коммуникации по истечении срока займа, преднастроенные предложения); (4) сегментация клиентской базы для адресных предложений и

приоритизации низкорисковых сегментов. Совокупно эти инструменты направлены на удвоение агентской сети (с 10 до 20–25 агентов) и увеличение доли повторных выдач без пропорционального роста стоимости привлечения.

Вместе с тем концентрация привлечения в доверительных каналах несёт риск зависимости от ограниченного круга агентов. Для его снижения целесообразны диверсификация агентской базы, документирование клиентских отношений в CRM (чтобы знание о клиенте принадлежало организации, а не агенту) и постепенное развитие собственных цифровых каналов как дополнения, но не замены доверительной модели.

Заключение

Проведённый анализ показывает, что в условиях зрелого и маржинально сжатого ломбардного рынка Казахстана (466 ломбардов, совокупный портфель 481,3 млрд тг на 1 апреля 2026 года) основными инструментами роста ссудного портфеля автоломбарда выступают не ценовые параметры продукта, а агентская сеть и доверительные каналы продаж. На примере ТОО «Ломбард Прогресс» установлено, что около 80 % сделок формируется через агентов и повторные обращения, обеспечивая одновременно низкую стоимость привлечения и поддержание качества портфеля. CRM-система является необходимым условием масштабирования этих каналов. Предложенный комплекс инструментов — мотивация агентов, реферальная программа, автоматизация повторных продаж и сегментация — увязан с целевыми показателями роста портфеля до 1,5 млрд тг при ROE не ниже 30 % и NPL не выше 3 %. Полученные выводы имеют практическую значимость для микрокредитных организаций ломбардного сегмента и могут быть распространены на сопоставимые рынки.

Список литературы:

1. Национальный Банк Республики Казахстан. Сведения о ломбардах Республики Казахстан по состоянию на 1 января и 1 апреля 2026 года. Астана, 2026. URL: https://www.nationalbank.kz (дата обращения: 22.06.2026).

2. О микрофинансовой деятельности: Закон Республики Казахстан от 26 ноября 2012 года № 56-V (с изм. и доп., внесёнными Законом РК от 16 января 2026 года № 259-VIII).

3. Ansoff H.I. Strategies for Diversification // Harvard Business Review. 1957. Vol. 35, № 5. P. 113–124.

4. Котлер Ф., Келлер К.Л. Маркетинг менеджмент. 15-е изд. СПб.: Питер, 2018. 848 с.

5. Reichheld F.F. The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value. Boston: Harvard Business School Press, 1996. 323 p.

6. Портер М. Конкурентная стратегия: методика анализа отраслей и конкурентов. М.: Альпина Паблишер, 2016. 454 с.

УДК 339.13:336.77

Ахметова Айсулу Калимтаевна

магистрант ОП «Деловое администрирование» (MBA)
Научный руководитель: Умирзаков Самажан Ынтыкбаевич
д.э.н., профессор, академик РАМ
Бизнес-школа, НАО «Университет Нархоз»
(г. Алматы,Казахстан)

РАЗРАБОТКА МОДЕЛИ УПРАВЛЕНИЯ ТЕХНОЛОГИЧЕСКИМ ПРЕОБРАЗОВАНИЕМ ТОРГОВОГО ПРЕДПРИЯТИЯ В УСЛОВИЯХ ЦИФРОВИЗАЦИИ (НА ПРИМЕРЕ РЫНКА «АЛТЫН ОРДА», РЕСПУБЛИКА КАЗАХСТАН)

Аннотация: В статье рассматривается модель управления технологическим преобразованием торгового предприятия в условиях цифровизации, разработанная на примере крупнейшего оптово-розничного агропродовольственного рынка Республики Казахстан «Алтын Орда». В качестве ключевого механизма реализации модели рассматривается комплекс инструментов финансовых технологий (FinTech). На теоретическом уровне формируется интегрированная концептуальная рамка, объединяющая теорию платформенной экономики, теорию бизнес-экосистем и прикладную теорию FinTech. На эмпирическом уровне выполнен анализ текущей бизнес-модели объекта (AS-IS), бенчмаркинг международных кейсов (Taobao Villages, Hepsiburada, Rungis), а также авторское полевое исследование (n = 45). Предложена целевая модель «Altyn Orda Digital Hub» (TO-BE), включающая четыре взаимосвязанных FinTech-инструмента: единый платёжный хаб, цифровой факторинг, альтернативный кредитный скоринг и автоматизированный учёт логистики. Прогнозный расчёт экономической эффективности на трёхлетнем горизонте подтверждает окупаемость проекта в течение 1,5 лет при объёме инвестиций 250 млн тенге, чистой приведённой стоимости порядка +275 млн тенге и внутренней норме доходности 38–42 %.

Ключевые слова: цифровая трансформация, FinTech, оптово-розничные рынки, платёжные экосистемы, цифровой факторинг, управление изменениями, рынок «Алтын Орда», Казахстан.

Введение

Современная экономика Республики Казахстан проходит этап масштабной цифровой трансформации финансового сектора. По данным Национального Банка Республики Казахстан, доля безналичных платежей в общем объёме операций возросла с уровня менее 30 % в 2015–2016 годах до 87,2 % к концу 2025 года; в 2024 году число платежей через QR-код впервые превысило число платежей через POS-терминалы (на 26 %), а ежедневный объём операций через платёжные системы НБРК в первом полугодии 2025 года составил около 6,3 трлн тенге. Доля «ненаблюдаемой» экономики, по данным Бюро национальной статистики Республики Казахстан, сократилась с 19,75 % ВВП в 2021 году до 16,71 % ВВП по итогам 2024 года.

Однако в сегменте крупных оптово-розничных рынков и продовольственных хабов уровень цифровизации остаётся существенно ниже среднестатистических показателей по экономике. Подобные торговые объекты исторически функционируют в условиях преобладания наличных расчётов, фрагментированной информационной среды и низкого уровня формализации деловых отношений. Указанные особенности не только препятствуют интеграции данных объектов в современную финансовую систему, но и создают системные риски: снижение собираемости налогов, ограничение доступа малого и среднего бизнеса (МСБ) к финансированию, рост издержек логистики и дефицит достоверной рыночной аналитики.

Объектом настоящего исследования выступает рынок «Алтын Орда» (Карасайский район Алматинской области) — крупнейший оптово-розничный агропродовольственный хаб Республики Казахстан. По данным Asian Development Bank (Allen, 2018), объект обеспечивает приблизительно 90 % национальной горизонтальной торговли в Казахстане при пропускной способности до 2 млн тонн плодоовощной продукции в год; на территории функционирует около 5 000 торговых точек, ежедневный пассажиропоток составляет приблизительно 50 000 человек.

Целью статьи является разработка модели управления технологическим преобразованием торгового предприятия в условиях цифровизации (на примере рынка «Алтын Орда») и обоснование её экономической эффективности. В качестве ключевого инструментального механизма реализации модели рассматривается интегрированный пакет инструментов финансовых технологий (FinTech). Научная новизна исследования заключается в формировании интегрированной концептуальной рамки, объединяющей теорию платформенной экономики, теорию бизнес-экосистем и прикладную теорию FinTech применительно к специфическому объекту — крупному агропродовольственному рынку развивающейся экономики, для которого подобная синтезирующая модель ранее не применялась. Практическая значимость состоит в формировании готового к внедрению комплекса управленческих решений, адаптированных к национальному контексту Республики Казахстан.

Теоретические основы исследования

Платформенная экономика и теория бизнес-экосистем

Каноническая теоретическая рамка платформенной экономики была заложена в работе Rochet и Tirole (2003), показавшей, что ценообразование на платформах должно учитывать перекрёстные внешние эффекты между сторонами: каждый дополнительный продавец повышает ценность платформы для покупателей, и наоборот. Van Alstyne, Parker и Choudary (2016) развили эту логику, доказав, что традиционные «трубопроводные» компании, к которым относятся и классические оптовые рынки, систематически проигрывают платформам в условиях зрелой цифровой экономики, поскольку платформы обладают принципиально лучшей предельной экономикой и сетевыми эффектами.

Концепция бизнес-экосистем, развиваемая в работах Adner (2021), описывает условия успешного формирования многоучастниковых деловых сред. Ключевое положение теории состоит в том, что устойчивость экосистемы определяется не количеством участников, а согласованностью их ценностных предложений и равновесностью распределения создаваемой ценности. Применительно к крупному оптово-розничному рынку платформенная логика означает переход от модели

«арендодатель — арендатор» к модели «оператор цифровой инфраструктуры — участники экосистемы».

FinTech как инструмент трансформации торговых экосистем

Финансовые технологии (FinTech) в современном понимании представляют собой не просто совокупность инструментов оплаты, а комплекс цифровых сервисов, обеспечивающих формирование доверия, ускорение оборачиваемости капитала и снижение информационной асимметрии между участниками деловых отношений (Arner, Buckley, & Zetzsche, 2020; Pazarbasioglu et al., 2020). По данным Всемирного банка (Pazarbasioglu et al., 2020) и Банка международных расчётов (Claessens et al., 2021), интеграция FinTech-сервисов в торговые потоки формирует так называемый «контур данные–сеть–активность» (DNA loop), в рамках которого транзакционные данные становятся источником финансовых продуктов, а финансовые продукты, в свою очередь, расширяют пользовательскую базу платформы.

Принципиально значимым результатом цифровой трансформации торговых экосистем является внедрение альтернативного кредитного скоринга на основе транзакционных данных. Berg, Burg, Gombović и Puri (2020) показали, что даже минимальный «цифровой след», формируемый при онлайн-регистрации и совершении первых транзакций, обладает прогностической силой по умолчанию, сопоставимой или превосходящей классические кредитные рейтинги. Hong Kong Monetary Authority (2021) и Consultative Group to Assist the Poor (CGAP, 2024) подтвердили жизнеспособность подобной модели на эмпирических данных индийских и гонконгских финтех-операторов.

Для оптовой торговли продовольствием особое значение имеет цифровой факторинг. По данным Всемирного банка (2021) и OECD (2021), цифровой факторинг сокращает срок оплаты поставок в среднем на 70 % и устраняет основной фактор кассовых разрывов в продовольственных цепочках МСБ.

Методология

Методологический аппарат исследования построен на сочетании кабинетных и полевых методов. Кабинетный блок включает контент-анализ нормативных актов Республики Казахстан, отчётно-аналитических материалов Национального Банка РК и Бюро национальной статистики РК, отчётов международных организаций (World Bank, BIS, OECD, McKinsey & Company), научных публикаций в Scopus и Web of Science, а также аналитических обзоров рынка платежей PwC Kazakhstan и Mastercard Kazakhstan.

Стратегический анализ макросреды и внутренней среды объекта выполнен с применением методов PESTEL и SWOT. Моделирование текущих и целевых бизнес-процессов осуществлено в логике BPMN-нотации (модели AS-IS и TO-BE). Бенчмаркинг международного опыта выполнен по трём кейсам, репрезентирующим различные географические и экономические контексты: Taobao Villages в Китайской Народной Республике, экосистемы Hepsiburada и Trendyol в Турции, оптовый рынок Rungis во Франции.

Полевая часть исследования включает анкетный опрос арендаторов рынка (n = 45), репрезентирующих основные товарные группы (плодоовощная продукция, сухофрукты и орехи, бакалея, мясная и молочная продукция, непродовольственные товары) и типы субъектов (индивидуальные предприниматели — 71,1 %; малые юридические лица — 28,9 %), а также серию из четырёх глубинных полуструктурированных интервью с

ключевыми экспертами рынка: представителем администрации, руководителем оптового направления крупного арендатора, представителем фермерского хозяйства, сотрудником банка-партнёра. Прогнозный расчёт экономической эффективности выполнен с применением классических методов инвестиционного анализа (NPV, IRR, ROI, анализ чувствительности).

Результаты исследования

Диагностика объекта: модель AS-IS и системные барьеры

Моделирование текущего состояния рынка «Алтын Орда» позволило идентифицировать четыре функциональных контура (сбор арендной платы, розничные расчёты, оптовые расчёты с поставщиками, логистическое обеспечение), в каждом из которых преобладает наличный расчёт. Совокупная доля наличного оборота в денежных потоках объекта, по экспертной оценке, составляет не менее 70 %. По итогам диагностики сформулированы пять системных барьеров технологической трансформации: информационная фрагментация, доминирование наличных расчётов, кассовые разрывы у поставщиков (отсрочки оплаты 7–45 дней), ограниченный доступ субъектов МСБ к финансированию и непрозрачность логистики.

Бенчмаркинг международных кейсов

Анализ трёх международных кейсов выявил три устойчивые архитектурные модели цифровизации крупных торговых хабов. Программа Taobao Villages (запущена Alibaba Group в 2014 году) объединяет в 2022 году 7 780 «таобао-деревень» и 2 429 «таобао-городов» в 28 провинциях, обеспечивает 8,28 миллиона рабочих мест в сельской местности и достигла объёма продаж 1,3 трлн юаней в 2021 году (AliResearch, 2020; Luo & Niu, 2019). Цифровой банк MYbank, входящий в группу Ant Financial, обслуживал свыше 20 миллионов сельских клиентов с уровнем просроченной задолженности около 1,5 % (MYbank, 2021).

Турецкий опыт характеризуется параллельным развитием двух доминирующих финтех-экосистем: Trendyol (250 тысяч активных продавцов, GMV 12,5 млрд долларов США в 2024 году) и Hepsiburada (полная платёжная лицензия, 15 миллионов пользователей кошелька Hepsipay, объём кредитования 16,2 млрд лир в 2024 году). Стратегически значимым для казахстанского контекста стало приобретение Hepsiburada казахстанской финтех-группой Kaspi.kz в январе 2025 года за 1,127 млрд долларов США, что подтверждает применимость модели супераппа в национальном контексте.

Французский кейс Rungis демонстрирует модель единого инфраструктурного оператора (SEMMARIS) с двадцатилетним инвестиционным планом цифровизации объёмом около 1,8 млрд евро. Совокупный товарооборот объекта составляет около 10 млрд евро в год, доля безналичных расчётов превышает 95 %, на B2B-платформе rungismarket.com консолидирован ассортимент 50+ оптовых продавцов в 120 товарных категориях.

Эмпирическое исследование участников рынка

Анкетный опрос подтвердил высокую готовность арендаторов к технологическим изменениям при наличии экономических стимулов. Ключевые результаты: 82,2 % респондентов принимают преимущественно наличные средства; 77,8 % готовы перейти на полностью безналичный расчёт при наличии материального стимула (скидка по аренде или доступ к недорогому кредиту); 73,3 % испытывают системную потребность в дополнительном оборотном капитале не реже одного раза в квартал; 60,0 % готовы

предоставить банку доступ к данным о товарообороте в обмен на упрощённый доступ к кредитному продукту. Для 35,6 % опрошенных переход на цифровой учёт ассоциируется с рисками налогового характера, что определяет необходимость целенаправленной работы по управлению изменениями.

Серия глубинных интервью раскрыла мотивационные основания выявленных закономерностей. Представитель администрации рынка подтвердил инвестиции около 600 млн тенге в первичную модернизацию инфраструктуры со стороны новой управляющей компании «Big City Market» (запуск официального сайта altynordaresmi.kz, автоматизированная оплата въезда по QR-коду, размещение POS-терминалов в павильонах, развёртывание около 300 камер видеонаблюдения). Представители оптового арендатора, фермерского хозяйства и банка-партнёра указали на отсутствие единой инфраструктуры на уровне всего объекта как ключевой барьер для интеграции цифровых решений.

Целевая модель «Altyn Orda Digital Hub» и оценка экономической эффективности

На основе результатов диагностики и бенчмаркинга разработана целевая модель «Altyn Orda Digital Hub» (TO-BE), архитектура которой построена на пяти принципах: единая инфраструктурная платформа, открытость к множественным финансовым партнёрам, модульность, равновесность распределения ценности, регуляторная встроенность. Стратегическим ядром модели является интегрированный пакет из четырёх FinTech-инструментов: единого платёжного хаба на базе QR-эквайринга и SoftPOS, цифрового факторинга, альтернативного кредитного скоринга и автоматизированного учёта логистики (RFID/распознавание госномеров).

Реализация модели предусмотрена в рамках двенадцатимесячной дорожной карты, разделённой на четыре последовательных этапа продолжительностью три месяца каждый. Логистический контур модели усилен интеграцией с реальным эскизным архитектурным проектом складского комплекса с кросс-докингами в селе Иргели Карасайского района Алматинской области, земельный участок 7 800 м², площадь застройки 1 658,91 м², помещение кросс-дока 1 234 м²).

Прогнозный расчёт экономической эффективности на трёхлетнем горизонте показывает, что чистый кумулятивный экономический эффект проекта переходит в положительную зону к началу 18-го месяца его реализации. Чистая приведённая стоимость (NPV) при ставке дисконтирования 15 % составляет порядка +275 млн тенге к концу третьего года; внутренняя норма доходности (IRR) — 38–42 % годовых; показатель рентабельности инвестиций (ROI) на трёхлетнем горизонте — около 118 %. Анализ чувствительности подтверждает экономическую устойчивость проекта: даже при одновременном наступлении трёх ключевых негативных факторов (замедление адаптации арендаторов на 25 %, рост капитальных затрат на 20 %, снижение комиссионных доходов на 30 %) NPV остаётся положительным.

Обсуждение

Результаты исследования позволяют сформулировать четыре управленчески значимых вывода. Во-первых, технологическое преобразование крупного оптово-розничного рынка целесообразно реализовывать в рамках платформенной модели управления, в которой управляющая компания выполняет роль оператора цифровой экосистемы, а не только арендодателя физического пространства. Подобный переход

меняет структуру монетизации объекта, диверсифицирует источники дохода и снижает зависимость от конъюнктуры арендных ставок.

Во-вторых, сложившийся макроэкономический и технологический контекст Республики Казахстан особенно благоприятен для предлагаемой модели: при доле безналичных платежей 87,2 % и доминировании QR-эквайринга над классическим POS-эквайрингом задача распространить эту модель на крупный оптово-розничный объект перестаёт быть пионерной. Технологическая база уже сформирована, а казахстанский потребитель обладает выраженной пользовательской привычкой к мобильным платежам.

В-третьих, наиболее релевантной для условий рынка «Алтын Орда» является комбинация архитектурных образцов: модели единого инфраструктурного оператора (по образцу Rungis), сквозной связки «торговля + платежи + кредитный сервис» (по образцу Taobao Villages) и собственного финтех-контура (по образцу Hepsiburada). Дополнительным контекстуальным преимуществом является то, что модель супераппа уже верифицирована на казахстанском рынке группой Kaspi.kz.

В-четвёртых, ключевым фактором успеха трансформации выступает сбалансированная стратегия управления изменениями. Эмпирическое исследование подтвердило, что основным мотиватором перехода арендаторов на цифровые сервисы является не сама технология, а связанные с ней экономические преимущества — прежде всего доступ к недорогому оборотному капиталу. Это определяет приоритетность экономической мотивации участников и необходимость формирования специализированного консультационного центра в составе администрации рынка.

Заключение

Разработанная управленческая модель «Altyn Orda Digital Hub» формирует целостный управленческий ответ на структурные барьеры цифровой трансформации крупных оптово-розничных рынков в условиях развивающейся экономики. Интегрированная концептуальная рамка, объединяющая теорию платформенной экономики, теорию бизнес-экосистем и прикладную теорию FinTech, обеспечивает теоретическое основание модели; результаты бенчмаркинга трёх международных кейсов и эмпирического исследования 45 арендаторов формируют доказательную базу её применимости в национальном контексте.

Прогнозный расчёт экономической эффективности подтверждает финансовую устойчивость проекта: при объёме инвестиций 250 млн тенге в первом году срок окупаемости составляет около 18 месяцев, NPV — порядка +275 млн тенге, IRR — 38–42 % годовых. Совокупный эффект внедрения модели охватывает три ключевые группы стейкхолдеров: для управляющей компании — рост и диверсификацию доходов, для арендаторов и оптовых поставщиков — доступ к недорогому оборотному капиталу и ликвидацию кассовых разрывов, для государства — снижение доли теневого оборота и рост налоговых поступлений.

Перспективы дальнейшего исследования связаны с разработкой методики количественной оценки социального эффекта проекта, расширением модели на другие крупные торговые объекты Республики Казахстан и интеграцией с региональными платёжными и логистическими экосистемами в контексте евразийского экономического сотрудничества.

Список использованных источников:

1. Adner, R. (2021). Winning the right game: How to play, compete, and succeed in the new world of ecosystems. MIT Press.
2. Allen, S. (2018). Kazakhstan national wholesale market master plans and wholesale distribution center development strategy. Asian Development Bank, Almaty-Bishkek Economic Corridor Project.
3. AliResearch. (2020). China Taobao Village research report 2020. Alibaba Group.
4. Arner, D. W., Buckley, R. P., & Zetzsche, D. A. (2020). FinTech, RegTech, and the reconceptualization of financial regulation. Northwestern Journal of International Law & Business, 37(3), 371–413.
5. Berg, T., Burg, V., Gombović, A., & Puri, M. (2020). On the rise of FinTechs: Credit scoring using digital footprints. The Review of Financial Studies, 33(7), 2845–2897. https://doi.org/10.1093/rfs/hhx134
6. Claessens, S., Frost, J., Turner, G., & Zhu, F. (2021). FinTech and the digital transformation of financial services: Implications for market structure and public policy. BIS Papers No. 117. Bank for International Settlements.
7. Consultative Group to Assist the Poor. (2024). Leveraging transactional data for micro and small enterprise lending. CGAP.
8. Frost, J., Gambacorta, L., Huang, Y., Shin, H. S., & Zbinden, P. (2019). BigTech and the changing structure of financial intermediation. BIS Working Papers No. 779. Bank for International Settlements.
9. Hepsiburada. (2024). Fourth quarter and full year 2024 financial results.
10. Hong Kong Monetary Authority. (2021). Alternative credit scoring of micro-, small and medium-sized enterprises. HKMA.
11. Kaspi.kz JSC. (2025). 4Q and FY 2024 financial results. NASDAQ: KSPI.
12. Luo, X., & Niu, C. (2019). E-commerce participation and household income growth in Taobao Villages. World Bank Policy Research Working Paper No. 8811.
13. McKinsey & Company. (2023). The 2023 McKinsey global payments report.
14. MYbank. (2021). MYbank's use of digital technology leads to record growth in rural clients [Press release]. BusinessWire.
15. OECD. (2021). Trade finance for SMEs in the digital era. OECD Publishing. https://doi.org/10.1787/e505fe39-en
16. Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). Platform revolution: How networked markets are transforming the economy. W. W. Norton & Company.
17. Pazarbasioglu, C., Garcia Mora, A., Uttamchandani, M., Natarajan, H., Feyen, E., & Saal, M. (2020). Digital financial services. World Bank Group.
18. Rochet, J.-C., & Tirole, J. (2003). Platform competition in two-sided markets. Journal of the European Economic Association, 1(4), 990–1029.
19. Uppler. (2022). Rungis (B2B marketplace): How to shift from a physical to a digital market. Case study.
20. Van Alstyne, M. W., Parker, G. G., & Choudary, S. P. (2016). Pipelines, platforms, and the new rules of strategy. Harvard Business Review, 94(4), 54–62.
21. World Bank. (2021). Technology and digitization in supply chain finance handbook. World Bank Group.

22. Бюро национальной статистики Агентства по стратегическому планированию и реформам Республики Казахстан. (2025). Объём ненаблюдаемой экономики Республики Казахстан.

23. Каргабаева, С. Т., Болаткызы, С., Камали, К. М., & Ахметова, К. И. (2025). Влияние цифровизации на электронную коммерцию в Казахстане: макроэкономические эффекты и прогноз. Экономика и статистика, 3, 53–67.

24. Национальный Банк Республики Казахстан. (2025). Аналитический обзор развития платёжных систем Республики Казахстан за первое полугодие 2025 года.

25. PwC Kazakhstan / Strategy&. (2024). Анализ рынка платежей в Республике Казахстан за 2023 год. PwC Kazakhstan.

26. Ulysmedia.kz. (2024). «Алтын Орда» по-новому: как меняется крупнейший рынок Казахстана.

UDC 06.00.00

Kali Ansar

11 M, Student
Nazarbayev Intellectual School of Science and Mathematics
in Medeu District of Almaty
(Almaty, Kazakhstan)

IMPACT OF INFLATION ON SMALL AND MEDIUM ENTERPRISES/HOW INFLATION AFFECTS SMALL AND MEDIUM ENTERPRISES' RESILIENCE AND STABILITY?

Abstract: Inflation serves as a major economics issue, that significantly impact small and medium enterprises, employment rates and market stability. This study investigates how inflation affects SMEs' resilience and stability in Kazakhstan. This research consists of primary data(survey, interviews and case studies) and secondary data from sources such as National Bank of Kazakhstan, Bureau of National Statistics. The results show that SMEs are getting pressed by inflation and in result, suffer from rising operational costs and reduced consumers' buying power. Although many businesses demonstrate adaptability through pricing adjustments and cost-control strategies, these responses do not always fully offset declining profitability. This study suggests that implication of such practical measures like financial planning, governmental support and access to affordable financing may significantly strengthen overall resilience and stability of SMEs. This research aims to find an answers to broad macroeconomic challenge from local businesses' responses.

Key words: Inflation, Small and Medium enterprises, business, enterpreneurship, purchasing power, economic uncertainity, business resilience, business stability, financial planning.

INTRODUCTION

While official economic indicators may present moderate inflation figures, many small and medium-sized enterprises experience much stronger pressure in their daily operations. Business survival mainly depends on supplier prices, increasing rent, wages and consumers' buying powerv – so even a small shift in any of thta attributes will create massively serious challeges for enterprises. This contrast highlights the importance of examining inflation not only as a macroeconomic statistic, but also as a real business experience.

Inflation is destructive for business activity and market confidence. The World Bank states that limited access to finance combined with inflationary pressure remains one of the key barriers to SME growth [11]. Rising costs are reducing profits, while uncertainty is holding back investment and growth. Moreover, the decline in electricity purchases weakens household demand, which directly affects small and medium enterprises that depend on domestic consumers.

This situation can be clearly seen in Kazakhstan, as SMEs play a crucial role here. They are affecting the employement creation, tax revenue and regional development. However, rising prices of goods, logistics, services, and public utilities put additional pressure on company owners. Although major corporations are affected, small and medium-sized companies often have fewer reserves and less capacity to manage customer complaints.

The objective of this study is to examine how inflation in Kazakhstan affects local businessses' resilience and stability and what practical measures they can implement in order to avoid it. To achieve this aim, study examines the major cost pressures faced by SMEs, changes in consumer demand, current adaptation strategies, and the role of financial or government support.

The central question of this study is : How Inflation affects Small and Medium Enetrprises' Resilience and Stability?

The hypothesis of this research is that if SMEs apply effective financial planning, adaptive pricing strategies, and receive targeted institutional support, they can significantly reduce the negative impact of inflation and improve long-term sustainability.

The object of this research is small and medium enterprises and local business that are operating in Kazakhstan within changing economic conditions. The subject of the research is the relationship between inflation, business performance, resilience, and operational stability.

The scientific novelty of this study lies in connecting micro and macroeconomics data with the real situations SME owners have to face. This research connects surveys, interviews, and case study for more practical understanding of inflation at business level, while previous studies mainly were focusing on national indicators.

This research may be useful for policymakers designing SME support programs, for enterpreneurs seeking strategies to regulate inflation and for society in understanding how economic instabillity contributes to the national backbone of national economy.

LITERATURE REVIEW

In recent times there has been growing attention on the economic factors affecting small and medium enterprises (SMEs) as a result of increasing inflation, worldwide instability, and changing policy measures. SMEs are broadly acknowledged as drivers of job creation and economic expansion, though experts and organizations question their capacity to endure extended periods of price volatility. The World Bank states that restricted access to finance, along with inflationary pressures, continues to be one of the main obstacles to the growth of SMEs globally [11]. Media outlets likewise portray inflation as an issue that limits both consumer spending and business viability [5]. Meanwhile, local statistics indicate that the effects of inflation on SMEs vary depending on state assistance, market composition, and access to financial resources, highlighting the need for a more nuanced assessment.

Multiple studies highlight the resilience of SMEs in the face of economic challenges. The European Commission reports that SMEs across the European Union sustained job growth during periods of significant inflation, demonstrating their structural adaptability and continued importance to the labor market [2]. Similarly, the European Investment Fund notes that public financial instruments reduce credit constraints and enhance SMEs' access to capital, particularly in times of instability [8].

Comparable trends appear in Kazakhstan. According to the Government of Kazakhstan, recent updates to business support policies aim to strengthen SMEs and ease financial burdens [12]. Likewise, the Bureau of National Statistics of Kazakhstan reports an increase in the number of registered SMEs, emphasizing their growing role in the national economy [4]. These findings collectively suggest that targeted government action can partially mitigate inflation-related risks and sustain SME operations.

Conversely, other sources emphasize the limitations imposed on SMEs by inflation. According to the National Bank of Kazakhstan, persistent inflation reduces purchasing power

while simultaneously driving up production and operating costs [3]. The Bureau of National Statistics of Kazakhstan confirms that rising prices for goods and services place additional pressure on businesses, particularly those with limited capacity to adjust their own prices [10].

News reports reflect these concerns. Tengrinews notes that rising inflation in Kazakhstan has weakened consumer demand, directly reducing SME revenues [6]. Similarly, LSM.kz reports that real household incomes have not kept pace with inflation, decreasing overall spending power [7]. On a global scale, the U.S. Chamber of Commerce observes declining business-owner confidence, attributing it to inflation, labor shortages, and price volatility [1]. Together, these findings indicate that inflation can negatively affect SME profitability when not sufficiently counterbalanced by policy interventions.

Overall, the literature presents a multifaceted view of SME performance in inflationary environments. While SMEs continue to contribute to employment and economic activity, inflation introduces challenges related to rising expenses, weakened consumer demand, and constrained investment capacity. Current research suggests that inflation is neither entirely harmful nor simple to manage; its effects depend on access to financing, consumer purchasing power, and the quality of state support. Rather than portraying SMEs as uniformly resilient or vulnerable, studies emphasize a conditional relationship shaped by economic policy and financial structures. Therefore, future research should focus on evaluating the long-term effectiveness of support mechanisms and balancing market pressures with government intervention.

METHODS

To answer the research questions, three research methods were used: a survey, interviews, and a case study. These methods were selected in order to collect both quantitative and qualitative data about how inflation affects small and medium enterprises (SMEs) in Kazakhstan and how these businesses maintain resilience and stability.

The survey method was used to collect numerical data from SME representatives. It allowed information to be gathered about rising operational costs, pricing adjustments, revenue changes, and financial challenges during inflation. This method provided measurable data on common problems faced by businesses.

The interview method was used to obtain detailed personal experiences from business owners. Through interviews, coping strategies, decision-making processes, and financial difficulties during inflation were explored.

The case study method was used to analyze inflation in a real business environment. A supermarket was selected to examine how inflation affects supply chains, pricing strategies, customer demand, and operational stability.

Survey participants included owners, managers, and employees of medium enterprises in Kazakhstan. Only individuals involved in business decision-making were included to ensure reliable and relevant data.

Two interviews were conducted:

• Aimanov Kairat – former owner of NomadXI store, which closed due to rising product prices caused by inflation.

• Imangalieva Kunsulu – owner of Masterlux Beauty Salon, who regularly adjusts service prices due to increasing costs of beauty products.

The case study focused on Esakhmet Supermarket, operating in Kazakhstan, where frequent price adjustments occur due to inflation.

Before data collection, permission was obtained from all participants. The purpose of the research was clearly explained, and participants were informed that their responses would be used only for academic purposes. Participation was voluntary. Participants were informed of their right to withdraw at any time. Confidentiality was respected where requested, and no sensitive financial information was published without approval. Ethical academic standards were followed throughout the research process.

Google Forms was used to design and distribute the survey, which received 75 responses. Responses were automatically collected and organized digitally.

Interviews were recorded using a voice recorder, and written notes were taken to ensure accuracy. Messaging applications were used to arrange meetings.

For the case study, observation notes, pricing comparisons, publicly available economic data, and interview insights were analyzed to understand operational changes caused by inflation.

The survey was conducted online using Google Forms to allow flexibility for business participants.

Interviews were conducted either online or face-to-face in quiet settings to ensure detailed responses.

The case study was conducted in a supermarket environment in Kazakhstan to observe pricing adjustments and supply management strategies in real conditions.

Interview 1

• Participant: Aimanov Kairat

• Former Owner of NomadXI Store

Introduction: Good afternoon. My name is Ansar. I am conducting research on how inflation affects small and medium enterprises in Kazakhstan. I would like to ask you several questions about your experience. Your answers will be used only for academic purposes.

• Question 1: How did inflation affect your business?

• Answer: Inflation seriously affected my store. Supplier prices increased every month, and I could not maintain stable prices for customers.

• Question 2: What was the biggest challenge?

• Answer: The biggest problem was unpredictability. I never knew how much the next delivery would cost, which made planning impossible.

• Question 3: Did you increase prices?

• Answer: Yes, but each increase reduced customer demand. Some customers stopped coming completely.

• Question 4: Why was the business closed?

• Answer: Expenses became higher than revenue. Rent, utilities, and product costs kept rising while sales declined.

• Question 5: What do SMEs need to survive inflation?

• Answer: Financial reserves are essential. Without savings or possible government support, survival becomes very difficult.

Interview 2

• Participant: Imangalieva Kunsulu

• Owner of Masterlux Beauty Salon

Introduction: Good afternoon. My name is Kali Ansar. I am researching how inflation impacts small and medium businesses in Kazakhstan. I would like to ask about your experience managing a beauty salon during inflation.

• Question 1: How does inflation affect your salon?

• Answer: Product prices such as shampoos and cosmetics increase regularly, which directly raises my operating costs.

• Question 2: How do you manage rising costs?

• Answer: I sometimes purchase products in bulk, but eventually service prices must be increased.

• Question 3: How do clients react?

• Answer: Some understand the situation, but others visit less often or choose cheaper services.

• Question 4: Has your profit changed?

• Answer: Profit margins have decreased because not all cost increases can be transferred to clients.

• Question 5: What helps your business remain stable?

• Answer: Customer loyalty and service quality are very important. Strong relationships help retain clients despite higher prices.

Case Study: Esakhmet Supermarket

The case study focused on Esakhmet Supermarket in Kazakhstan.

It was observed that product prices are adjusted frequently due to supplier price increases. Imported goods are more affected by inflation than locally produced items.

Several strategies have been implemented:

• Gradual price adjustments

• Increased cooperation with local suppliers

• Promotion of discounted essential goods

Changes in customer behavior were observed. Consumers increasingly prioritize essential goods over luxury products.

It was concluded that although inflation creates operational pressure, larger SMEs such as supermarkets demonstrate greater resilience due to diversified suppliers and higher sales volume.

RESULTS

The results of this research were obtained by combining three research methods: a survey, two interviews with business owners, and a case study of a supermarket operating in Kazakhstan. The purpose of combining these methods was to obtain both quantitative and qualitative data in order to understand how inflation affects small and medium enterprises (SMEs) and financial decision-making. The survey provided numerical data about public perceptions of inflation and its influence on daily expenses and business activity. Interviews with business owners helped identify real operational challenges faced by SMEs during inflation. Finally, the case study allowed the research to observe how inflation influences pricing strategies and business operations in a real business environment.

Survey Results

The survey collected responses from more than fifty participants. The demographic results showed that the largest group of respondents were individuals under 18 years old (35.8%), followed by participants aged 18–25 (26.4%). Smaller proportions included

individuals aged 26–40 (13.2%), 41–65 (13.2%), and over 65 years old (11.3%). In terms of gender distribution, 54.9% of respondents were male and 45.1% were female.

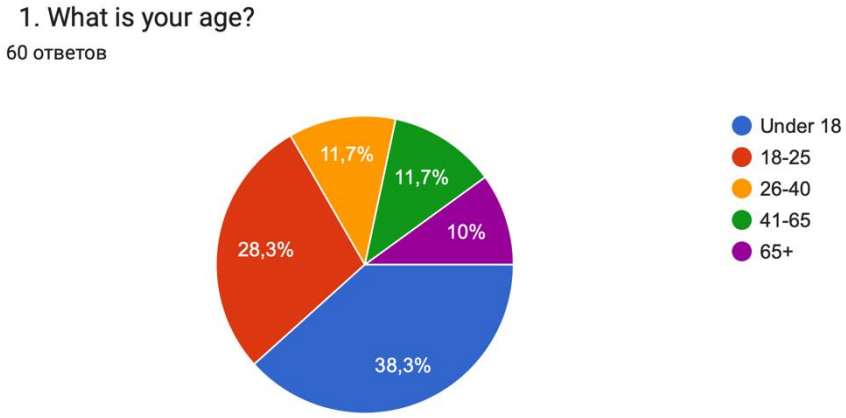


Figure 1. Demographic Results

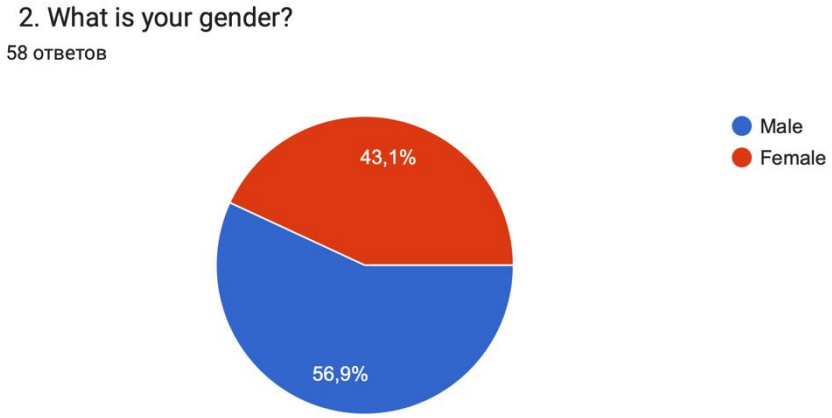


Figure 2. Gender Distribution

Regarding employment status, 41.5% of respondents identified themselves as students, 22.6% were employees, 20.8% were self-employed, and 15.1% were business owners. This distribution allowed the research to capture perspectives from individuals with different economic roles.

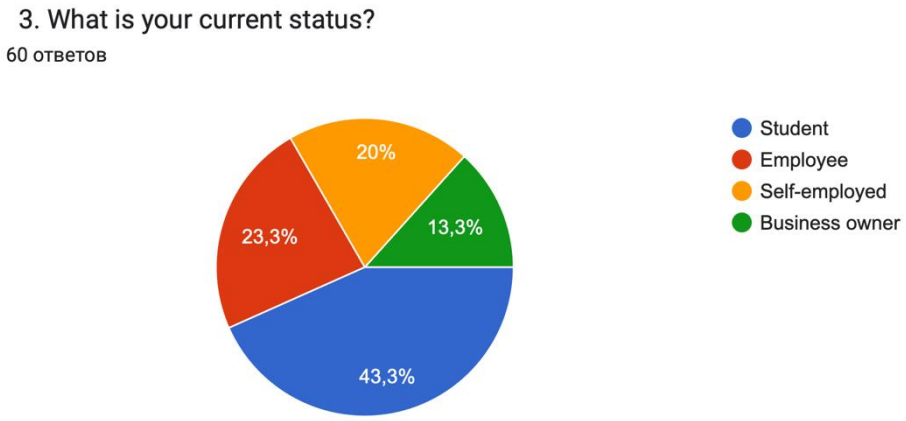


Figure 3. Employment Status

Participants were also asked how strongly inflation affected their daily expenses. The results showed that inflation significantly influences personal financial behavior. Approximately 39.6% of respondents reported that inflation affects their expenses very strongly. In addition, 20.8% reported strong effects and another 20.8% reported moderate effects. Only 11.3% reported slight influence and 7.5% stated that inflation had no impact on their daily spending.

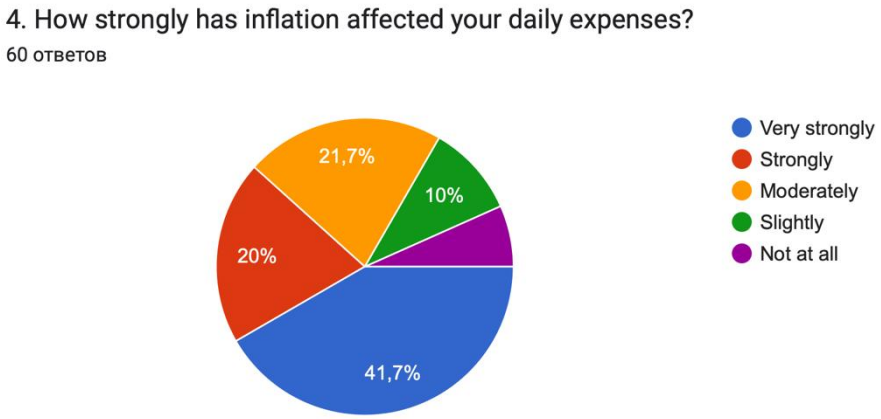


Figure 4. Inflations' Effect

The survey also explored attitudes toward small and medium enterprises. The results indicated that 52.8% of respondents believe SMEs play an important role in society and the economy. However, 20.8% disagreed and 26.4% were uncertain.

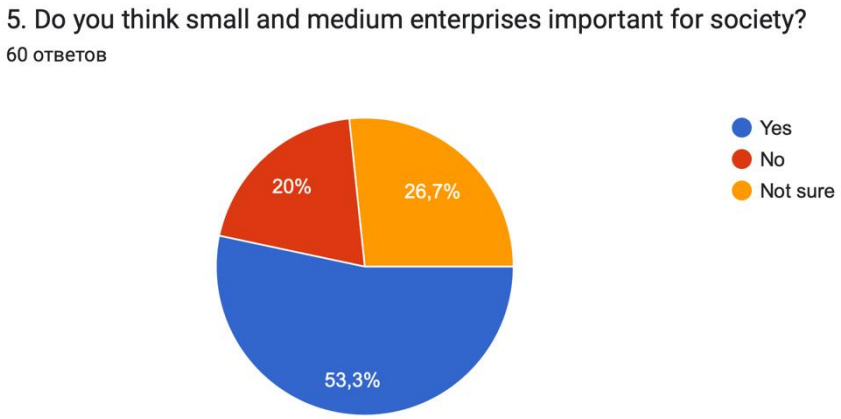


Figure 5. Importance of SMEs in Society

Another question asked whether businesses require government support during periods of high inflation. The majority of respondents (57.7%) agreed that government assistance is necessary to help businesses survive economic instability. Meanwhile, 17.3% disagreed and 25% were unsure.

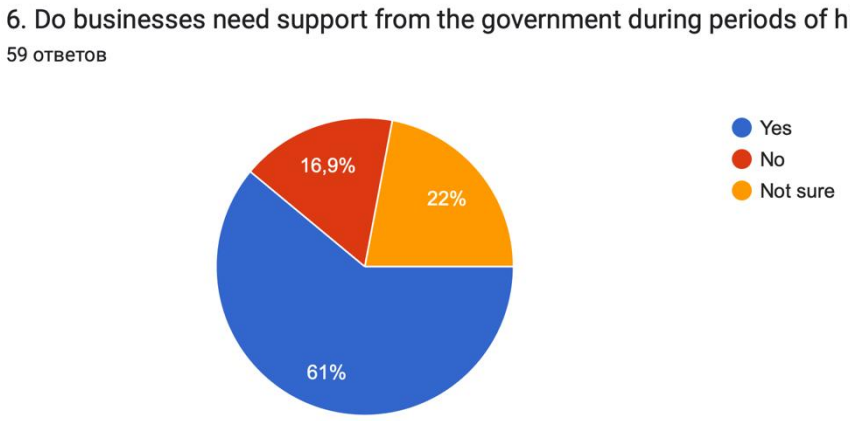


Figure 6. Government's Intervention

Finally, participants were asked whether small businesses should adopt financial planning tools during inflation. A strong majority of respondents (64.7%) supported the adoption of new financial planning strategies to help businesses manage rising costs and economic uncertainty. Only 15.7% disagreed while 19.6% were uncertain.

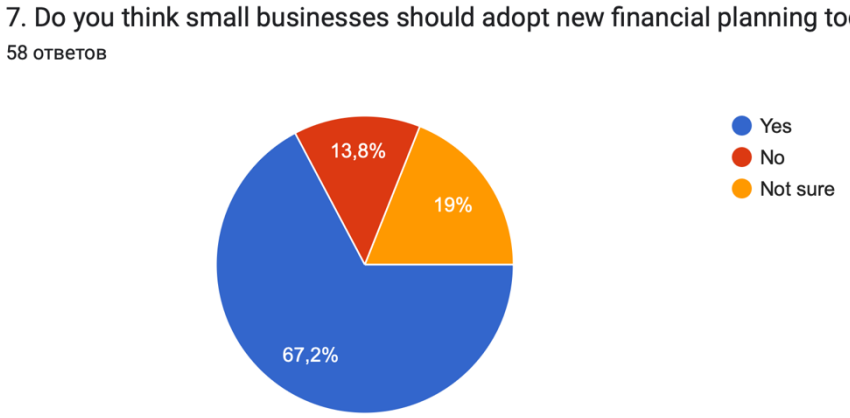


Figure 7. Adaption to New Financial Planning Tools

Interview Results

The interviews provided deeper insight into the experiences of business owners dealing with inflation. The first interview was conducted with Aimanov Kairat, the former owner of NomadXI store. He explained that inflation caused continuous increases in supplier prices, which made it impossible to maintain stable product pricing. As costs continued to increase, customer demand decreased. Eventually, operational expenses exceeded revenue, forcing the business to close.

The second interview was conducted with Imangalieva Kunsulu, the owner of Masterlux Beauty Salon. According to the interview, inflation increased the cost of professional cosmetics and hair products used in the salon. To manage these rising costs, the business owner occasionally purchased supplies in bulk and gradually increased service prices. However, profit margins still decreased because not all cost increases could be transferred to customers. Both interviews demonstrate that inflation creates financial pressure on SMEs, particularly through increased supply costs and unstable demand.

Case Study Results

The case study focused on Esakhmet Supermarket. Observations showed that inflation significantly influenced product pricing and supply management. The supermarket frequently adjusted prices due to increases in supplier costs, particularly for imported goods.

Several adaptation strategies were observed. These included gradual price adjustments, cooperation with local suppliers to reduce dependence on imports, and promotional discounts for essential goods. In addition, changes in consumer behavior were noticed. Customers increasingly prioritized essential goods such as food products while reducing purchases of non-essential items.

Overall, the case study indicates that larger SMEs such as supermarkets demonstrate greater resilience during inflation because they have diversified suppliers, higher sales volumes, and more flexible pricing strategies.

DISCUSSION

This study helps to understand how inflation affect small and medium enterprises' resilience and stability, where resilience is understood by ability of adapting for changing economics factors. Findings show that inflation has a enormous and complex effect on business' overall situation, significantly affecting on its demand, costs, pricing.

This results show that SME's stabillty may weaken by inflation, because it is affecting on operational costs, logistics and it also decreasing consumer's purchasing power. As observed in both case study and survey, consumers decide to prioritize on essential goods and purchase less non-essential things. Rising prices created a shift in behaviour of consumers, suggesting that SMEs, particularly those operating in discretionary sectors, experience less revenue in inflationary periods. This findings are correspondent and consistent with existing literature, which underscores that inflation erodes income and deteriorate consumers' demand.

In addition, results suggest that rising scores represent a major constraint on small and medium enterprises. Interviews taken from SMEs' owners show that supplier prices massively rise during inflation, putting additional pressure on businesses to adjust their prices and pricing strategies. But, SMEs may not be able to pass these increased costs on consumers without losing demand. This reflects the dynamics of cost-push inflation, where higher production and operational expenses reduce revenue and make financial management harder. As a result, businesses face both decline in consumers' demand and profitability of business.

Another important aspect is that inflation bring uncertainity with itself, which negstively affect SME's resilience. Effectiveness of long-term planning get significantly shaped and shifted by inability to predict future costs. Instead, businesses may rely on short-term decision making which may reduce efficiency and increase financial risk. Given results align with previous research, offering that economics instability and uncertainity discourages investment, strategic and financial planning, particularly for smaller businesses with limited finances.

Despite these obstacles, findings suggest that SMEs show a certain level of resilience through adaptive strategies. Adjusting prices gradually, purchase goods in bulk or shift to other more profitable supplies – all that shows that SMEs can react to changes and somehow manage costs. All that shows that SMEs are capable of adapting to economics instability and changes, underlining and supporting definition of resilience used in this study. Also, this is consistent with literature that underscores the flexibility of small and medium enterprises compared to large firms. But, results suggest that such strategies may only partially mitigate the effects of inflation, as a considerable number of businesses continue to experience reduced profit margins.

Also, resilience and stability are not evenly distributed across SMEs, which is shown in this study. More established and experienced larger business appear to be more capable of adapting of inflation due to greater access to financial opportunities. For example, smaller businesses with stricted financial opportunities may be more sensibile and vulnerable to economics shocks, ups and downs. This observation supports existing research highlighting that access to finance and institutional support programs plays essential role in determining SMEs performance during periods of economics instabllity.

These findings have several important implications. First of all, SMEs may benefit from improved financial planning and cost management strategies to be aware of inflationary pressures. Second of all, adaptability is key factor in maintaining operations, because businesses that can answer immediately to changing conditions have higher 'survival' rate. Third of all, findings prove that government policies and financial support programs can be crucial in stabilizing SMEs and reducing negative impact of inflation.

Despite that, such limitations like survey size and etc. have to be considered, because they are deteriorating quality of research and authencity of findings. Survey sample size may not fully represent the broader population, which is also limits generalizibillity of research. Also, qualitative data collected is based on small number of interviews(n=2), which not capture a diversity of different types of SMEs, which makes findings not applicable for some businesses. Case study which focuses on single business may not reflect wider industry trends.

Future research could expand this study by adding more interview with the owners of different types of SMEs because inflation affects differently on businesses. By including larger and more diverse sample of SMEs across multiple industries and regions. Quantative financial data could be used to provide more precise insights of how inflation affects stability and profitability.

CONCLUSION AND FURTHER RESEARCH

The findings of this study demonstrate that inflation have a significant impact on both resilience and stability of SMEs. By combining survey data , interviews and case study, this research show how inflation affects business overall performance.

The results suggest that inflation reduces SME's stability by rising operational costs while weakening consumer's demand. Its putting business in challenging environment, where business must operate, and in result leading to low revenues and reduced profitabilllity. At the same time, inflation brings uncertainity which is limiting the ability of SMEs to engage in effective long-term planning.

Overall, this study suggests that inflation creates a complex and challenging context for SMEs, affecting both their ability to remain stable and to adapt to changing conditions. While some businesses demonstrate resilience, many remain vulnerable due to rising costs, reduced demand, and economic uncertainty. Therefore, addressing the impact of inflation on SMEs is essential for supporting economic stability and ensuring sustainable business development in the long term.

References:

1. U.S. Chamber of Commerce. Small Business Index – Q4 2024 Quarterly Spotlight. 2024. URL: https://www.uschamber.com/sbindex/2024-Q4/quarterly-spotlight

2. European Commission. SME Performance Review 2024: SMEs driving job creation despite high inflation. 2024. URL: https://single-market-economy.ec.europa.eu/news/new-

2024-smeperformance-review-smes-driving-job-creation-despite-high-inflation-2024-07-04_en

3. National Bank of Kazakhstan. Official inflation and economic indicators. 2024. URL: https://nationalbank.kz/en/page/inflyacionnye-ozhidaniya

4. Bureau of National Statistics of Kazakhstan. Business statistics publication. 2024. URL: https://stat.gov.kz/en/industries/economy/stat-kon-obs/publications/288738/

5. BBC News. Global inflation and economic pressures. 2024. URL: https://www.bbc.com/news/articles/c75k002qkzno

6. Tengrinews. Inflation in Kazakhstan has accelerated. 2024. URL: https://en.tengrinews.kz/kazakhstan_news/inflation-in-kazakhstan-has-accelerated-266569/

7. LSM.kz. Purchasing power of Kazakhstan's population. 2024. URL: https://lsm.kz/pokupatel-skaya-sposobnost-kazahstancev

8. European Investment Fund. EIF Working Paper 2024/101. 2024. URL: https://ideas.repec.org/p/zbw/eifwps/313620.html

9. Atameken. SMEs in numbers: Income, taxes, and development. 2024. URL: https://stat.gov.kz/en/industries/business-statistics/stat-org/publications/223654/

10. Bureau of National Statistics of Kazakhstan. Price statistics publication. 2024. URL: https://stat.gov.kz/en/industries/economy/prices/publications/282971/

11. World Bank. SME finance. 2024. URL: https://www.worldbank.org/ext/en/topic/competitiveness/small-and-medium-enterprises-smes-finance

12. Government of Kazakhstan. Revising business support measures. 2024. URL: https://primeminister.kz/en/news/implementation-of-presidents-address-government-revises-business-support-measures-28950

УДК 339.138:004.738.5

Камал Айару Жумаханкызы

Магистрант 2 курса
Научный руководитель: Еспергенова Ляззат Рыскулбековна,
к.э.н., Associate Professor
Школа менеджмента
Алматы Менеджмент Университет
(г. Алматы, Казахстан)

ТРАНСФОРМАЦИЯ МАРКЕТИНГОВЫХ СТРАТЕГИЙ В ЭЛЕКТРОННОЙ КОММЕРЦИИ КАЗАХСТАНА: ПОВЕДЕНЧЕСКИЙ АСПЕКТ И ЭКОСИСТЕМНЫЙ ПОДХОД

Аннотация: В статье исследуется трансформация маркетинговых стратегий на рынке электронной коммерции Республики Казахстан в условиях его активного развития и повышения уровня цифровой зрелости. Показано, что по мере формирования устойчивых моделей онлайн-потребления традиционные подходы, ориентированные преимущественно на привлечение новых покупателей, постепенно уступают место стратегиям удержания клиентов (Customer Retention). На основе анализа статистических данных Бюро национальной статистики Республики Казахстан и аналитических отчетов PwC Kazakhstan рассмотрены изменения в потребительском поведении, в частности снижение среднего чека, обусловленное распространением практики регулярных онлайн-покупок товаров повседневного спроса. Особое внимание уделено влиянию национальных финтех-экосистем, а также культурно-языковых факторов на формирование лояльности потребителей и повышение пожизненной ценности клиента (Lifetime Value, LTV). Сделан вывод о том, что устойчивое развитие предприятий электронной коммерции связано с переходом от модели разовых продаж к построению долгосрочных отношений с потребителями.

Ключевые слова: электронная коммерция, потребительское поведение, программы лояльности, удержание клиентов, финтех-экосистемы, локализация, маркетплейсы.

Современный этап развития электронной коммерции в Республике Казахстан характеризуется качественной трансформацией. Из категории инновационного формата совершения покупок онлайн-торговля окончательно перешла в разряд привычной модели потребительского поведения для большинства населения страны. По данным Бюро национальной статистики Агентства по стратегическому планированию и реформам Республики Казахстан, уровень цифровой вовлеченности населения достиг исторически высокого уровня: более 96 % граждан являются активными пользователями сети Интернет [1]. Это создало благоприятные условия для дальнейшего развития электронной коммерции.

Согласно совместному аналитическому отчету PwC Kazakhstan и Ассоциации цифрового Казахстана (DKA), по итогам 2024 года объем розничной электронной торговли в стране достиг 3 439 млрд тенге (около 7,3 млрд долларов США), что на 42 %

превышает показатель предыдущего года [2]. В первом полугодии 2025 года положительная динамика сохранилась: объем рынка составил 1,7 трлн тенге, увеличившись еще на 19 % по сравнению с аналогичным периодом предыдущего года [3]. Доля электронной коммерции в общем объеме розничного товарооборота достигла 17,1 %, постепенно приближаясь к среднемировому уровню, который оценивается в 20,5 % [3].

Вместе с тем интенсивный рост рынка сопровождается усилением конкурентной среды. В этих условиях экстенсивная модель развития, основанная преимущественно на постоянном привлечении новой аудитории, постепенно утрачивает эффективность. Увеличение стоимости цифровой рекламы и рост расходов на привлечение покупателей снижают рентабельность первичных продаж. В связи с этим все большее значение приобретают маркетинговые стратегии, ориентированные не только на привлечение новых клиентов, но и на их удержание, формирование лояльности и развитие долгосрочных отношений с потребителями.

Теоретические основы удержания потребителей в цифровой среде

В научной литературе электронный маркетинг традиционно рассматривался с позиций транзакционного подхода, в рамках которого основным результатом маркетинговой деятельности считался факт совершения покупки [4]. Однако развитие цифровой экономики и изменение моделей потребительского поведения обусловили переход к концепции маркетинга взаимоотношений, ориентированной на формирование долгосрочного взаимодействия с клиентом и управление его жизненным циклом (*Customer Lifecycle Management*).

Одними из ключевых показателей эффективности в рамках данной концепции являются стоимость привлечения клиента (*Customer Acquisition Cost* – CAC) и пожизненная ценность клиента (*Lifetime Value* – LTV). Показатель CAC отражает совокупные затраты компании на маркетинговые мероприятия, рекламу и стимулирование продаж, необходимые для привлечения нового клиента и совершения им первой покупки [4]. В свою очередь показатель LTV характеризует совокупный доход или прибыль, которую компания получает от клиента за весь период взаимодействия с ним [5].

В современной маркетинговой практике одним из ориентиров эффективности считается соотношение LTV и CAC на уровне не ниже 3:1, свидетельствующее о способности бизнеса компенсировать затраты на привлечение клиентов за счет их долгосрочной ценности [5]. Вместе с тем Т. В. Панкина, А. Ф. Никишин и А. В. Бойкова отмечают, что стремление компаний к снижению затрат на привлечение клиентов без формирования эффективной системы их последующего удержания приводит к росту показателя оттока (*Churn Rate*) и снижению долгосрочной устойчивости бизнеса [6].

Значительное внимание вопросам формирования потребительской лояльности уделяется и в отечественных исследованиях. Так, Н. Е. Толеубаева отмечает, что внедрение комплексных программ лояльности в цифровых B2C-сервисах способствует увеличению показателя LTV на 10–15 % и одновременному снижению коэффициента оттока клиентов на 4–6 процентных пунктов [7]. Полученные результаты подтверждают экономическую эффективность перераспределения маркетинговых ресурсов от преимущественного привлечения новых клиентов к развитию механизмов их удержания.

Специфика потребительского поведения на казахстанском рынке e-commerce

Анализ особенностей потребительского поведения на рынке электронной коммерции Республики Казахстан позволяет выделить ряд устойчивых тенденций, оказывающих существенное влияние на формирование современных маркетинговых стратегий.

1. Феномен «микропокупок» и падение среднего чека. Согласно аналитическим материалам PwC Kazakhstan, в 2024 году на рынке электронной коммерции Казахстана наблюдалась разнонаправленная динамика основных показателей. Несмотря на увеличение количества онлайн-транзакций на 80 % (до 91 млн операций во втором полугодии), средний чек сократился на 24 % и составил 21,2 тыс. тенге [2]. Снижение среднего чека обусловлено постепенной интеграцией электронной коммерции в повседневную жизнь населения. Онлайн-покупки перестали восприниматься исключительно как способ приобретения дорогостоящих товаров и все чаще используются для регулярного приобретения продукции повседневного спроса [8]. Существенную роль в этом процессе сыграло развитие сети пунктов выдачи заказов, обеспечившее удобный доступ к товарам и снизившее барьеры для совершения небольших покупок. В результате потребители стали чаще приобретать товары категорий FMCG (продукты питания, бытовая химия, косметика и др.), однако объем отдельных заказов уменьшился [8].

Полученные данные согласуются со статистикой Бюро национальной статистики Республики Казахстан за 2024 год. Средний чек при покупке товаров через маркетплейсы составил 17 572 тенге, тогда как при оформлении заказов на собственных интернет-ресурсах брендов данный показатель достиг 39 179 тенге [9]. Аналогичная закономерность наблюдается и на рынке услуг: средний чек на маркетплейсах составил 3 274 тенге, а на специализированных интернет-площадках – 15 243 тенге [9]. Это позволяет сделать вывод, что маркетплейсы преимущественно используются для приобретения товаров повседневного спроса, тогда как фирменные интернет-магазины сохраняют свои позиции в сегментах, предполагающих более высокий уровень вовлеченности потребителя в процесс выбора товара.

2. Роль маркетплейсов в развитии электронной коммерции Казахстана. Одной из ключевых особенностей рынка электронной коммерции Казахстана является высокая концентрация продаж на маркетплейсах. Согласно данным PwC Kazakhstan, их доля в общем объеме онлайн-продаж в денежном выражении достигла 91 %, а в структуре всех совершаемых транзакций – 94 % [2]. Это свидетельствует о том, что маркетплейсы выступают основными каналами взаимодействия продавцов и покупателей, концентрируя значительную часть потребительского спроса.

Использование традиционных каналов привлечения трафика, включая контекстную рекламу в поисковых системах, постепенно уступает место инструментам *Retail Media*, предполагающим размещение рекламных материалов непосредственно внутри торговых платформ на этапе выбора товара потребителем [10]. Как отмечает Р. Саядян, внутренние рекламные инструменты маркетплейсов, включая продвижение товарных карточек, баннерную рекламу и приоритетное размещение в поисковой выдаче платформы, демонстрируют более высокую конверсию по сравнению с внешними рекламными каналами, поскольку ориентированы на аудиторию, уже находящуюся на этапе принятия решения о покупке [10].

3. Культурно-языковой фактор как драйвер лояльности. В условиях усиления конкуренции одним из факторов повышения потребительской лояльности становится локализация клиентского опыта. Адаптация цифровых сервисов с учетом языковых и культурных особенностей целевой аудитории способствует укреплению доверия к бренду и формированию долгосрочных отношений с потребителями.

Согласно результатам исследования Yandex Ads, не менее 15 % интернет-пользователей Казахстана предпочитают получать информацию о товарах и услугах на казахском языке [11]. Проведенный анализ показывает, что аудитория, взаимодействующая с локализованным казахоязычным контентом, характеризуется более высоким уровнем вовлеченности и доверия к брендам, что положительно влияет на показатели удержания клиентов [11]. Реализация мультиязычной SEO-стратегии позволяет компаниям расширять охват аудитории и повышать эффективность поискового продвижения. В частности, продвижение по казахоязычным запросам характеризуется более низким уровнем конкуренции по сравнению с русскоязычным сегментом, что обеспечивает дополнительные возможности для повышения видимости бренда в поисковых системах.

4. Экосистемный маркетинг и финансовые стимулы. Одной из отличительных особенностей казахстанского рынка электронной коммерции является тесная интеграция торговых платформ с крупнейшими финтех-экосистемами, ведущие позиции среди которых занимают Kaspi.kz и Halyk. В этих условиях потребительская лояльность формируется не только по отношению к отдельному продавцу, но и к цифровой экосистеме, объединяющей платежные сервисы, маркетплейсы, финансовые продукты и государственные услуги [8].

Основные маркетинговые механизмы удержания клиентов, используемые крупнейшими финтех-экосистемами Казахстана, представлены в таблице 1.

Таблица 1 – Маркетинговые механизмы удержания клиентов в финтех-экосистемах РК [8].

Экосистема	Преимущества для удержания (Retention)	Недостатки и барьеры для выхода (Switching Costs)
Kaspi.kz	Удобство транзакций через Kaspi QR. Быстрый онбординг (выдача карты за 5 минут). Выгодные беспроцентные рассрочки. Интеграция с госуслугами внутри Super-App.	Ограничение операций рамками тенговых счетов. Отсутствие поддержки международных кошельков (Apple/Google Pay). Комиссии за внешние переводы.
Halyk	Развитая бонусная программа Halyk Bonus. Поддержка мультивалютности. Полноценная интеграция с Apple Pay и Google Pay. Бесшовный переход в Halyk Market и Halyk Travel.	Исторически сложный интерфейс (исправлено в последних обновлениях). Меньшая плотность эмоционального контакта с пользователем. Отсутствие интеграции государственных сервисов.

Представленные в таблице данные позволяют сделать вывод о различиях в подходах к формированию потребительской лояльности. Экосистема Kaspi.kz

ориентирована преимущественно на создание высокой степени вовлеченности пользователей за счет интеграции платежных сервисов, государственных услуг и финансовых инструментов в единую цифровую среду. Подобная модель повышает издержки переключения потребителей на альтернативные сервисы и способствует долгосрочному удержанию клиентов [8].

В отличие от Kaspi.kz, экосистема Halyk делает акцент на развитии открытой инфраструктуры и расширении возможностей использования бонусной программы Halyk Bonus. Поддержка мультивалютных операций, интеграция с международными платежными сервисами Apple Pay и Google Pay, а также взаимодействие с Halyk Market и Halyk Travel формируют дополнительные преимущества для пользователей и способствуют повышению их лояльности.

Существенным элементом современных маркетинговых стратегий выступает использование инструментов потребительского кредитования и программ рассрочки. Возможность приобретения товаров в рассрочку сроком до 24 месяцев при установленном кредитном лимите до 6 млн тенге снижает финансовый барьер для покупателей и способствует увеличению числа повторных покупок, особенно в сегменте дорогостоящих товаров [5].

Заключение. Проведенное исследование позволяет сделать вывод о том, что маркетинговые стратегии в электронной коммерции Республики Казахстан меняются под влиянием цифровизации экономики, трансформации потребительского поведения и усиления конкуренции. Рост доли электронной торговли в структуре розничного товарооборота сопровождается переходом от модели, ориентированной преимущественно на привлечение новых покупателей, к стратегиям удержания клиентов и формирования долгосрочной потребительской лояльности.

Анализ статистических данных показал, что снижение среднего чека при одновременном увеличении количества онлайн-покупок отражает изменение потребительской модели. Электронная коммерция все чаще используется для регулярного приобретения товаров повседневного спроса, что повышает значимость повторных покупок и удержания клиентов.

Развитие электронной коммерции в Казахстане во многом определяется распространением маркетплейсов и финтех-экосистем, объединяющих торговые, платежные и сервисные функции. В этих условиях конкурентные преимущества компаний формируются не только за счет ценовой политики, но и благодаря совершенствованию клиентского опыта, развитию программ лояльности, персонализации предложений и использованию современных финансовых инструментов.

Важную роль для казахстанского рынка играет локализация цифровых коммуникаций с учетом языковых и культурных особенностей аудитории. Использование многоязычного контента способствует расширению охвата, укреплению доверия к бренду и повышению эффективности взаимодействия с потребителями.

Таким образом, устойчивое развитие предприятий электронной коммерции связано с использованием инструментов удержания клиентов, развитием экосистемного подхода и адаптацией маркетинговых стратегий к изменениям потребительского поведения.

Список литературы:

1. Кажиева Ж. Х., Қуантқан Б., Сапарова Б. С., Агумбаева А. Е. Тенденции и перспективы развития электронной коммерции в Республике Казахстан // Вестник университета «Туран». – 2024. – № 2. – С. 214–229.
2. Lim N., Abrossimova V., Kaniyeva N. Retail e-Commerce Market in the Republic of Kazakhstan: Analysis for 12 Months of 2024. – Almaty: Strategy&, PwC Kazakhstan, Digital Kazakhstan Association, 2025. – 25 p.
3. Lim N., Abrossimova V., Kaniyeva N. Retail e-Commerce Market in the Republic of Kazakhstan: Analysis for 6 Months of 2025. – Almaty: Strategy&, PwC Kazakhstan, Digital Kazakhstan Association, 2026. – 24 p.
4. Бурук А. Ф. Тенденции развития электронной коммерции и интернет-маркетинга // Интерэкспо Гео-Сибирь. – 2019. – Т. 5. – С. 183–188.
5. Kulman A. B., Yermekbayeva D. D., Celik S. Impact of Digital E-Commerce Platforms on Consumer Purchase Behavior: Management Perspectives and Service Perception // Bulletin of "Turan" University. – 2025. – № 4. – P. 192–204.
6. Панкина Т. В., Никишин А. Ф., Бойкова А. В. Привлечение и удержание покупателей в электронной торговле // Российское предпринимательство. – 2018. – Т. 19, № 3. – С. 783–792.
7. Толеубаева Н. Е. Влияние программ потребительской лояльности на удержание клиентов и рост LTV в цифровых сервисах // Вестник науки. – 2025. – № 12 (93). – Т. 3. – С. 317–325.
8. Об электронной коммерции в Республике Казахстан (2024 г.): статистический бюллетень / Бюро национальной статистики Агентства по стратегическому планированию и реформам Республики Казахстан. – Астана, 2025.
9. Рост потребления контента на казахском языке: экономические и маркетинговые аспекты // PayGate Kazakhstan. – URL: https://paygate.kz/ (дата обращения: 29.06.2026).
10. Саядян Р. Влияние внутренних и внешних маркетинговых инструментов на управление поведением потребителей на маркетплейсах // Практический маркетинг. – 2025. – № 4 (334). – С. 40–45.
11. Внутренние аналитические материалы / Исследование Yandex Ads и Clickshunter.kz. – 2025–2026.

ЖАРАТЫЛЫСТАНУ ҒЫЛЫМДАРЫ – ЕСТЕСТВЕННЫЕ НАУКИ – NATURAL SCIENCES

ӘОЖ 544.478.1:

Кеңес Әмина Серікқызы

2-курс магистратура студенті
Химиялық процестер және өнеркәсіптік экология кафедрасы
Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті
(Алматы қ., Қазақстан)

АРОМАТТЫ НИТРОҚОСЫЛЫСТАРДЫ ТИІМДІ ГИДРЛЕУГЕ АРНАЛҒАН АКРИЛАМИД-МЕТАКРИЛ ҚЫШҚЫЛЫ НЕГІЗІНДЕГІ КРИОГЕЛЬ-ИММОБИЛИЗАЦИЯЛАНҒАН АЛТЫН НАНОКАТАЛИЗАТОРЛАРЫ

Аңдатпа. Бұл мақалада ароматты нитроқосылыстарды селективті гидрлеуге арналған акриламид пен метакрил қышқылы (ААм-МАҚ) негізіндегі макрокеуекті криогельдерді синтездеу және оларға алтын нанобөлшектерін (AuNPs) иммобилизациялау нәтижелері ұсынылған. Зерттеу барысында алынған ААм-МАҚ/AuNPs каталитикалық жүйесі 4-нитрофенолды ағынды реакторда тотықсыздандыруда сыналып, жоғары каталитикалық белсенділік пен қайта қолдануға төзімділік көрсетті. Ұсынылған композиттік материалдар "жасыл химия" принциптеріне сай экологиялық қауіпсіз катализаторлар ретінде перспективалы болып табылады.

Кілт сөздер: Криогель, алтын нанобөлшектері, иммобилизация, ароматты нитроқосылыстар, 4-нитрофенол, ағынды реактор, макрокеуекті тасымалдағыш.

1. Кіріспе

Хош иісті (ароматты) нитроқосылыстарды сәйкес аминдерге дейін гидрлеу реакциясы химия өнеркәсібіндегі, органикалық синтездегі және фармацевтикадағы ең маңызды процестердің бірі болып табылады [1]. Түзілетін ароматты аминдер полимерлік материалдар, бояғыштар және биологиялық белсенді қосылыстар өндірісінде алмастырылмайтын аралық өнімдер рөлін атқарады.

Дәстүрлі ұнтақ тәрізді гетерогенді катализаторлар агрегацияға бейім келеді және оларды реакция аяқталғаннан кейін сұйық ортадан бөліп алу күрделі технологиялық шығындарды талап етеді. Бұл мәселені шешудің тиімді жолдарының бірі — нанобөлшектерді үш өлшемді (3D) макрокеуекті полимерлі матрицаларға, яғни криогельдерге иммобилизациялау [2].

Криогельдердің бірегей макрокеуекті құрылымы диффузиялық шектеулердің болмауын, ағынды жүйелерде төмен гидравликалық кедергіні және каталитикалық белсенді орталықтардың жоғары қолжетімділігін қамтамасыз етеді [3]. Осыған байланысты, бұл жұмыстың мақсаты акриламид пен метакрил қышқылы негізіндегі криогельдерді синтездеу, алтын нанобөлшектерін иммобилизациялау және олардың модельдік реакциядағы каталитикалық қасиеттерін зерттеу болып табылады.

2. Материалдар мен зерттеу әдістері

Криогель синтезі: ААм-МАҚ криогелі бос радикалды криополимерлену әдісімен синтезделді. Процесс барысында 450 мг акриламид (ААм) 9 мл суда ерітіліп, оған 343 мг метакрил қышқылы (МАҚ) қосылды. Торлағыш агент ретінде 25 мг N,N'-метиленбисакриламид (МБА) енгізілді. Реакциялық қоспа 0 °C-қа дейін салқындатылған соң, инициирлеуші жүйе ретінде 7 мкл ТЕМЕД және 0,83 мл 10% аммоний персульфаты ерітіндісі қосылды. Қоспа -12 °C температурада криотермостатта қатырылды.

Нанобөлшектерді иммобилизациялау: Жуылған және кептірілген криогельдер алтын нанобөлшектерімен (AuNPs) *in situ* әдісі арқылы иммобилизацияланды. Бұл үшін криогель үлгісі $HAuCl_4$ ерітіндісіне батырылып, жүйе қосымша тотықсыздандырғыштарсыз 70 °C температурада қыздырылды.

Морфологиялық зерттеулер: Криогельдердің құрылымы мен морфологиясы JSM-6390 LV сканерлеуші электрондық микроскопия (СЭМ) құрылғысында зерттелді.

Каталитикалық сынақтар: Алынған ААм-МАҚ/AuNPs катализаторының белсенділігі ағынды типтегі реакторда сыналды. Модельдік реакция ретінде 4-нитрофенолды (4-НФ) натрий боргидриді ($NaBH_4$) қатысында тотықсыздандыру үдерісі таңдалды. Реакция кинетикасы ультракүлгін-көрінетін спектроскопия (UV-Vis) әдісімен бақыланды [4].

3. Нәтижелер және оларды талқылау

3.1. Криогель мен катализатордың құрылымдық сипаттамалары

СЭМ әдісімен алынған микрофотографияларды талдау ААм-МАҚ криогельдерінің дамыған, өзара байланысқан макрокеуекті архитектураға ие екенін көрсетті. Кеуектер мен арналардың өлшемдері бірнеше ондағаннан бірнеше жүз микрометрге дейін өзгереді, бұл криополимерлену кезіндегі мұз кристалдарының шаблондық әсерін тікелей растайды.

$HAuCl_4$ ерітіндісінде қыздыру барысында түссіз криогельдің қызыл-қоңыр түске өзгеруі байқалды, бұл коллоидтық алтынға тән плазмондық жұтылу құбылысының қалыптасқанын дәлелдейді [5]. ААм-МАҚ матрицасының құрамындағы карбоксил және амид топтары алтын иондарын тотықсыздандыруға және түзілген нанобөлшектерді агрегациядан қорғап тұрақтандыруға белсенді қатысады. Кескін талдауы (ImageJ) бөлшектердің 50%-дан астамының диаметрі 100 нм-ден кіші екенін көрсетті, олар макрокеуектің қабырғаларына біркелкі бекітілген.

3.2. Каталитикалық белсенділікті бағалау

4-нитрофенолды тотықсыздандыру Габер (Haber) схемасы бойынша бірнеше аралық кезеңдерді қамтитын күрделі механизммен жүреді, алайда оның жалпылама иондық теңдеуі келесідей өрнектеледі:

$$R - NO_2 + 6H^+ + 6e^- \rightarrow R - NH_2 + 2H_2O$$

Ағынды типтегі каталитикалық реакторда жүргізілген тәжірибелер катализатордың жоғары тиімділігін көрсетті. Реакция ерітіндісі кеуекті матрица арқылы өткізілгендіктен, өнімді катализатордан бөлу кезеңі барынша жеңілдетілді.

Кинетикалық зерттеулер тотықсыздану жылдамдығының температураға тікелей тәуелділігін байқатты. Температура 25 °C-тан 45 °C-қа дейін көтерілгенде реакция уақыты едәуір қысқарды, бұл процестің кинетикалық үдеуімен сипатталады. Тұрақтылықты бағалау үшін жүйе 10 бірізді каталитикалық цикл бойы сыналды. Нәтижелер көрсеткендей, криогельге иммобилизацияланған алтын нанобөлшектері көп

мәрте қолданылғаннан кейін де өз белсенділігі мен құрылымдық тұтастығын жоғалтпай, жоғары конверсия деңгейін сақтап қалды [6].

4. Қорытынды

Жүргізілген ғылыми-зерттеу жұмысының нәтижесінде келесідей қорытындылар жасалды:

1. Акриламид және метакрил қышқылы негізіндегі макрокеуекті криогельдер алтын нанобөлшектерін иммобилизациялауға арналған тиімді тасымалдағыш ретінде сәтті синтезделді.
2. Криогельдің дамыған макрокеуекті архитектурасы мен химиялық функционалдығы алтын нанобөлшектерін *in situ* тотықсыздандыруға және тұрақтандыруға қолайлы орта қалыптастыратыны дәлелденді.
3. Алынған ААм-МАҚ/AuNPs жүйесі ароматты нитроқосылыстарды (4-нитрофенол) ағынды типтегі реакторда тотықсыздандыру барысында жоғары хемоселективтілік, белсенділік және тұрақтылық танытты.
4. Ұсынылып отырған нанокатализаторлар қалдықсыз және энергияны аз қажет ететін технологияларды дамытуда, сондай-ақ көп мәрте қолдануға жарамды өнеркәсіптік ағынды реакторларды жасауда зор практикалық маңызға ие.

Қолданылған әдебиеттер тізімі:

1. Corma A., Serna P. Chemoselective Hydrogenation of Nitro Compounds with Supported Gold Catalysts // Science. – 2006. – Vol. 313, No. 5785. – P. 332-334.

2. Kudaibergenov S. E. Monograph: Macromolecular Complexes of Precious Metals and Their Catalytic Activity // IPMT Publishing. – Almaty, 2021. – 210 p.

3. Kudaibergenov S., Adilov Z., Berillo D., Tatykhanova G., Sadakbaeva Z., Abdullin K., Galaev I. Novel macroporous amphoteric gels: Preparation and characterization // Express Polymer Letters. – 2012. – Vol. 6, № 5. – P. 346–353.

4. Klivenko A., Tatykhanova G., Nuraje N., Kudaibergenov S. Hydrogenation of p-nitrophenol by gold nanoparticles immobilized within macroporous amphoteric cryogel based on N,N-dimetylaminoethylmetacrylate and methacrylic acid // Вестник Карагандинского университета. Серия "Химия". – 2015. – Vol. 4, № 80. – P. 6-12.

5. Kudaibergenov S. E., Tatykhanova G. S., Selenova B. S. Polymer Protected and Gel Immobilized Gold and Silver Nanoparticles in Catalysis // Journal of Inorganic and Organometallic Polymers and Materials. – 2016. – Vol. 26. – P. 1-14.

6. Zhao P., Feng X., Huang D., Yang G., Astruc D. Basic concepts and recent advances in nitrophenol reduction by gold- and other transition metal nanoparticles // Coordination Chemistry Reviews. – 2015. – Vol. 287. – P. 114-136.

Электронный научный журнал «Central Asian Scientific Journal»

Редактор: Байдильдинов Т.Ж.
Комп.верстка: Хусаинов Е.М.

Электронный научный журнал «Central Asian Scientific Journal»
-2026-2(30)-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