

ПӘНДЕР АРАЛЫҚ ҒЫЛЫМДАР – МЕЖДИСЦИПЛИНАРНЫЕ НАУКИ – INTERDISCIPLINARY SCIENCES

UDC 81'272:004.738.5

Albiyev Asser

Grade 11 student
Quantum STEM School
(Astana, Kazakhstan)

HOW MUCH OF AN ARTICLE COUNT IS HUMAN WORK? DECOMPOSING THE CREATION OF FIVE WIKIPEDIAS, 2010–2025

Abstract: The number of articles in a language's Wikipedia is routinely quoted as evidence of that language's presence online, including in a 2026 study of Kazakh digital language policy that reports 241,858 articles for the Kazakh edition. An article count is a count of pages, not of human work, and pages can be created automatically. Using the Wikimedia Analytics API, this study decomposes every content page created in the Kazakh, Uzbek, Kyrgyz, Azerbaijani, Tatar and Russian editions between January 2010 and August 2025 by the class of account that created it. Accounts carrying the platform's bot flag created 50.4 per cent of the Kazakh pages, 64.1 per cent of Kyrgyz and 90.0 per cent of Tatar, against 0.3 per cent of Azerbaijani. The flag is then shown to be insufficient. A burst rule — a month in which human-flagged page creation is at least ten times its own trailing median — identifies four Kazakh months, thirteen Uzbek and ten Kyrgyz, and none at all in Azerbaijani, Tatar or Russian. In those months a newly created page receives 2.5 to 3.8 subsequent edits against 17.1 to 21.8 in ordinary months, and carries roughly half the net bytes, which is the signature of automated stub creation by accounts the flag does not catch. Treating the burst excess as automation lowers the human share of Kazakh creation from 48.3 to 27.2 per cent. The same rule run as a placebo on already-flagged bot creation recovers 87 to 97 per cent of it in the three editions where it fires, so the signature it detects is automation rather than volume. Ranking the five non-Russian editions by articles gives Tatar first and Azerbaijani fourth; ranking them by human-created pages reverses this, and Azerbaijani leads on both corrected and uncorrected counts. Two findings are specific to Kazakh: no flagged bot has created an article there since 2012, and the human share of edits rose from 17.2 to 95.1 per cent over the period, but this rise is a collapse in bot activity rather than growth in human activity, and the core editing community averaged 44.2 people a month during the 2011 campaign against 37.1 in 2024. An article count should not be read as a measure of human work without this decomposition, and the platform's own bot flag is not sufficient to perform it.

Keywords: Wikipedia; Kazakh language; bots; automated content; digital language vitality; Central Asia; article count

1. Introduction

Wikipedia keeps a public counter of how many articles each of its language editions contains, and that counter has become a shorthand for how well a language is doing online. The habit reaches into the research literature. A 2026 study of Kazakh digital linguistic sovereignty lists the Kazakh Wikipedia among its indicators of the language's digital presence, reporting

that "the section comprised 241,858 articles covering a wide range of subjects, with approximately 2 million edits and a membership of over 10,000 registered participants" [1]. The count is used there exactly as it is used in public argument: as a quantity of Kazakh-language knowledge that exists.

The counter does not measure that. It measures pages. A page can be written by a person over several sittings, or generated by a program from a database in a fraction of a second, and the counter treats the two identically. Where automated creation has been heavy this is not a small correction. The Cebuano edition of Wikipedia is the extreme case: it is the largest non-English edition by article count, and almost the whole of it was produced by a single article-creating program [11].

Kazakhstan has a specific reason to care. In 2011 the foundation WikiBilim ran a campaign to grow the Kazakh edition, and the printed Kazakh National Encyclopedia donated its content to it [12]; a Wikimedia Foundation post from July 2012 records that "Today we have more than 130,000 articles and more than 200 active editors" [12]. Fifteen years later that campaign is still inside the counter that gets quoted. Nobody has said how much of it.

This paper asks a narrow, answerable question: of the content pages created in a language edition between 2010 and 2025, what share was created by a human being? We ask it for Kazakh and for four comparison editions — Uzbek, Kyrgyz and Azerbaijani, which face broadly similar circumstances, and Tatar, which is known for heavy automated creation — with Russian as a large reference.

The first answer is the obvious one, and it is already substantial: on the platform's own bot flag, half of the Kazakh pages created since 2010 were created by bots. The second answer is the one we think matters more. The flag is not sufficient. Mass creation also runs through ordinary, unflagged accounts, and it leaves a measurable trace that the flag misses. Correcting for that trace roughly halves the human share again, and it changes which edition sits at the top of the table.

We report one further result that cuts against a simple pessimistic reading. Kazakh bot creation stopped in 2012 and has not resumed, and the human share of Kazakh editing has risen steadily since. We also show why that rise is weaker evidence than it looks.

2. Background and related work

Automated accounts have been part of Wikipedia almost since its beginning, and the research on them is mostly about what they do rather than how much of the encyclopedia they built. Geiger describes bots as socio-technical actors embedded in the community's rules rather than as mere tools [7]. Halfaker and Riedl distinguish bots proper from "cyborgs" — human editors working through semi-automated tools that combine computational speed with human judgement [6]. That distinction is central to this paper: a cyborg edits from a human account, and no bot flag is attached to it. Tsvetkova and colleagues show that bots interact with each other in patterned ways over long periods [5], and Zheng and colleagues build a taxonomy of bot functions on the English edition and classify 1,601 of them [8], establishing that "bot" is not one kind of thing.

The work closest to ours in method is Steiner's, which monitors real-time edit activity across all language editions and reports that Wikipedias and Wikidata together are edited roughly 50 per cent by bots and 23 per cent by anonymous users [3, 4]. Steiner measures edits. We measure page creation, which is the quantity the article counter actually accumulates, and we measure it per edition over fifteen years rather than in a live window.

Reeves and Simperl review the literature on automated content generation for Wikipedia [9]. It is a review of generation techniques and evaluation methods; it does not provide a per-edition accounting of how many existing articles were machine-made, which is the gap this paper addresses.

On the other side sits the literature that treats online presence as a measure of a language's health. Kornai argues that most of the world's languages will not make the transition to digital use at all, and that the digital domain is where language vitality is now decided [2]. That framing is what gives an article count its rhetorical force, and it is also what makes an uncorrected count dangerous: if the count is the vitality measure, then a language can appear to be thriving on the strength of work no speaker did.

To our knowledge no study decomposes the article creation of the Central Asian Turkic Wikipedia editions by the class of account responsible, and none tests whether the platform's own bot flag is adequate for such a decomposition. Both are what we do here.

3. Data

All series come from the Wikimedia Analytics REST API [10], which is public, requires no account and no key, and exposes monthly aggregates of the MediaWiki history. We use six editions — kk, uz, ky, az, tt and ru — over the 188 months from January 2010 to August 2025, the last month available at the time of the pull on 23 September 2026.

For each edition and each month we pulled four series, each broken out by the four editor classes the API distinguishes (`'user'`, `'group-bot'`, `'name-bot'`, `'anonymous'`): content-namespace edits, newly created content pages, distinct editors, and net bytes added. We also pulled human editors split by monthly activity level, and the live article count from each edition's own MediaWiki statistics endpoint. The API returns only months in which a series is non-zero; an absent month is read as zero, which matters for bot series that run in short campaigns.

Two properties of the classes should be stated plainly. `'group-bot'` means the account held the bot user group; `'name-bot'` means the account's name contains "bot" without that group; `'anonymous'` means an unregistered editor. The classification is a property of the account, not of the behaviour, and Section 5.2 is about what that costs.

The window covers most but not all of each edition's stock. Pages created in the window number 97 per cent of the live Kazakh article count, 99 per cent of Kyrgyz, 85 per cent of Azerbaijani, 81 per cent of Uzbek, 75 per cent of Russian and 70 per cent of Tatar. The residual is pages created before 2010 net of pages deleted since. We therefore state our primary results as a decomposition of the 2010–2025 creation flow, which is fully measured, and treat any statement about the present stock as an extrapolation that carries the coverage ratio with it.

4. Method

The first step is arithmetic: for each edition, sum created content pages over the window by editor class and express each class as a share.

The second step is the substantive one. If mass creation also runs through unflagged human accounts, the class shares understate automation, and the understatement will be worst exactly where the campaigns were largest. We need a way to detect automation from behaviour rather than from account status.

We use a burst rule. For each edition and each month from the thirteenth onward, compute the median of human-flagged page creation over the preceding twelve months. A month is a creation burst when human-flagged creation is at least $*K*$ times that trailing median and also clears an absolute floor of 300 pages, which prevents the rule from firing on the ordinary noise

of a very small edition. The burst excess is the month's creation minus its own trailing median — that is, the part that is not explained by the edition's own recent human pace. We treat the burst excess as automation. Because a genuine human surge would also be caught, the excess is an upper bound on unflagged automated creation, and the flagged-bot count is a lower bound; the truth lies between, and we report both.

We set $K^* = 10$ for the main results and sweep K^* over 5, 10, 20 and 50, and the floor over 100, 300 and 1,000, so that no conclusion rests on a chosen threshold.

A rule of this kind must be shown to detect automation rather than simply volume, so we subject it to three checks. First, two behavioural corroborators computed independently of the rule: the number of subsequent edits a newly created page receives, and the net bytes added per new page, both compared between burst and ordinary months. Automated stub creation should produce pages that are barely revisited and short. Second, a placebo: the identical rule applied to creation that is *already* flagged as bot. If the burst shape is the automation signature, it should recover most flagged-bot creation; if it merely flags large months, it proves nothing. Third, the editions where we expect no unflagged campaigns — Azerbaijani, Russian — function as negative controls, and the rule is expected to stay silent there.

Finally, for the question of whether the human share is growing, we aggregate to years and fit ordinary least squares of the annual human share of edits on the year, reporting the slope, its t^* statistic and the fit. Sixteen annual points is a small sample and we treat weak t^* statistics as weak evidence rather than as evidence of absence.

5. Results

5.1 What the article counts are made of

Table 1 — Composition of content-page creation, January 2010 to August 2025.

"Bot" is the platform's own flag (group-bot plus name-bot). Core human editors is the mean monthly count of accounts making 25 or more content edits, over the last twelve months of the window. Articles per core editor divides the live article count by that mean.

Edition	Live articles	Created 2010–2025	Coverage	Bot share of creation	Bot share of edits	Core human editors / month	Articles per core editor
Kazakh (kk)	245,273	238,743	97%	50.4%	49.9%	34	7,232
Uzbek (uz)	361,810	292,212	81%	38.7%	57.3%	58	6,211
Kyrgyz (ky)	76,549	75,853	99%	64.1%	36.7%	7	11,067
Azerbaijani (az)	217,531	185,357	85%	0.3%	25.1%	73	2,973
Tatar (tt)	709,127	499,864	70%	90.0%	87.1%	8	94,550
Russian (ru)	2,118,986	1,589,009	75%	3.6%	19.3%	1,071	1,979

Half of everything the Kazakh edition created since 2010 was created by flagged bots: 120,333 pages against 115,427 from human-flagged accounts and 2,983 from unregistered editors. Kyrgyz is higher at 64.1 per cent and Tatar higher still at 90.0 per cent. Uzbek, at 38.7 per cent, looks the healthiest of the four.

The outlier runs the other way. The Azerbaijani edition has 217,531 articles, within eleven per cent of Kazakh, and flagged bots created 533 of the 185,357 pages it added in the window — three tenths of one per cent. Two editions of comparable size, built in almost opposite ways.

The last column is the crudest possible sanity check and it is already informative. Tatar carries 94,550 live articles for every core human editor active in the last year. Kazakh carries 7,232, Azerbaijani 2,973, Russian 1,979. An encyclopedia is maintained, not just created, and these ratios describe how much material sits behind each person available to maintain it.

5.2 The flag is not enough

Table 1 rests entirely on account status. Section 2 gives the reason to distrust it: semi-automated editing from ordinary accounts is a recognised category [6], and it carries no flag.

The Kazakh series shows the problem immediately. Human-flagged accounts created a median of 271 pages a month across the window. In June 2011 they created 18,448, in July 21,972, in August 9,456. Those three months are 229, 243 and 63 times the trailing twelve-month median, and they alone account for 49,876 pages.

Table 2 — Burst detection on human-flagged page creation, $K = 10$, floor = 300 pages.

Burst excess is creation above the month's own trailing twelve-month median. The last two columns are computed independently of the rule.

Edition	Burst months	Burst excess pages	Excess as share of human creation	Human share, flag only	Human share, corrected	Later edits per new page, burst	Later edits per new page, ordinary
Kazakh (kk)	4	50,498	43.7%	48.3%	27.2%	2.8	17.1
Uzbek (uz)	13	92,530	52.1%	60.7%	29.1%	3.8	19.6
Kyrgyz (ky)	10	13,755	52.3%	34.7%	16.6%	2.5	21.8
Azerbaijani (az)	0	0	0.0%	97.8%	97.8%	—	21.9
Tatar (tt)	0	0	0.0%	9.7%	9.7%	—	10.9
Russian (ru)	0	0	0.0%	89.7%	89.7%	—	39.5

The rule fires in three editions and stays silent in three. Where it fires, it fires hard: the excess is 43.7 per cent of all human-flagged Kazakh creation, 52.1 per cent of Uzbek and 52.3 per cent of Kyrgyz. Treating that excess as automation cuts the human share of Kazakh creation from 48.3 to 27.2 per cent, of Uzbek from 60.7 to 29.1 and of Kyrgyz from 34.7 to 16.6.

The last two columns are the reason to believe this is not an artefact of the rule. They are not part of the rule and could have come out flat. In Kazakh burst months a newly created page goes on to receive 2.8 later edits; in ordinary Kazakh months, 17.1. The same contrast appears in Uzbek (3.8 against 19.6) and Kyrgyz (2.5 against 21.8). Pages created in burst months are made once and left. Net bytes point the same way, if less sharply: 2,950 bytes per new page in Kazakh burst months against 6,953 in ordinary ones, 2,870 against 7,679 in Kyrgyz, 3,745 against 4,738 in Uzbek.

The Uzbek bursts are recent and sustained — eight months across 2022, and June 2024 alone with 37,456 pages against a trailing median of 1,174. The Kazakh bursts are all in 2011,

and they coincide with the documented WikiBilim campaign and the Kazakh National Encyclopedia donation of that year [12]. We take that coincidence as corroboration that the rule detects a real event, not as an identification of who performed it, which the aggregate data cannot establish.

5.3 Does the detector point the right way?

Table 3 — Placebo. The identical burst rule applied to creation that already carries the bot flag.

A rule that detects automation should recover most of it; a rule that merely detects large months would not.

Edition	Pages created by flagged bots	Falling in months the rule flags	Share recovered
Kazakh (kk)	120,333	105,291	87%
Uzbek (uz)	113,174	51,456	45%
Kyrgyz (ky)	48,555	47,319	97%
Azerbaijani (az)	248	0	0%
Tatar (tt)	440,707	389,384	88%
Russian (ru)	15,121	7,161	47%

Run against creation we already know to be automated, the rule recovers 87 per cent of it in Kazakh, 97 per cent in Kyrgyz and 88 per cent in Tatar. It recovers nothing in Azerbaijani, where flagged bots created only 248 pages in fifteen years and never enough in one month to clear the floor — the correct answer, not a failure. Uzbek and Russian sit near half, meaning their flagged bot work is more evenly spread and less campaign-shaped; where that is so, our correction is conservative, since steady automation does not burst.

Taken with the two behavioural corroborators in Table 2, this is the evidence that the burst shape is the automation shape. It is not proof that every flagged month was automated, and we do not claim it: the excess remains an upper bound.

Table 4 — Corrected human share of content-page creation under different burst thresholds K, floor held at 300.

Varying the floor over 100, 300 and 1,000 pages moves Kazakh not at all (27.2% throughout), Uzbek from 29.0% to 29.8% and Kyrgyz from 16.0% to 20.2%.

K	Kazakh	Uzbek	Kyrgyz	Azerbaijani	Tatar	Russian
5	24.8%	24.8%	15.6%	87.4%	9.7%	89.7%
10	27.2%	29.1%	16.6%	97.8%	9.7%	89.7%
20	27.6%	38.4%	17.9%	97.8%	9.7%	89.7%
50	27.6%	56.8%	28.4%	97.8%	9.7%	89.7%

The sensitivity sweep separates a robust conclusion from a fragile one, and we report both. The Kazakh correction is robust: across a tenfold range of *K* the corrected human share moves only between 24.8 and 27.6 per cent, and changing the floor does not move it at all. The Uzbek correction is not: at *K* = 5 it gives 24.8 per cent and at *K* = 50 it gives 56.8 per cent. The Uzbek campaigns are long enough to raise their own trailing median, so how much of them the rule catches depends on where the threshold sits. We therefore claim a level for Kazakh and only a direction for Uzbek. Kyrgyz sits in between, ranging from 15.6 to 28.4 per cent, and Azerbaijani flips only at the loosest threshold, where a single 2020 month is caught.

5.4 The league table, three ways

Table 5 — Five editions ranked by the headline article count and by human-created pages in the 2010–2025 window, before and after the burst correction. Russian excluded as the reference case.

Rank	By article count	By human-created pages (flag only)	By human-created pages (corrected)
1	Tatar — 709,127	Azerbaijani — 181,308	Azerbaijani — 181,308
2	Uzbek — 361,810	Uzbek — 177,468	Uzbek — 84,938
3	Kazakh — 245,273	Kazakh — 115,427	Kazakh — 64,930
4	Azerbaijani — 217,531	Tatar — 48,525	Tatar — 48,525
5	Kyrgyz — 76,549	Kyrgyz — 26,312	Kyrgyz — 12,557

Ranking by the public counter puts Tatar first and Azerbaijani last but one. Ranking by pages a person created inverts that: Azerbaijani first, Tatar fourth, and the gap between first and fourth is a factor of 3.7 rather than the factor of 3.3 by which Tatar appeared to lead. Azerbaijani leads on both the uncorrected and the corrected count, and it is the only edition whose position is unchanged by the correction — because there is nothing there to correct.

For Kazakh the two counts differ by a factor of 3.8: 245,273 articles, and 64,930 pages created by a human being under the corrected reading — 115,427 if only the flag is trusted. Both figures are for the 2010–2025 window; the live stock includes pre-2010 pages, which for Kazakh is a small residual given the 97 per cent coverage.

5.5 Is the human share growing?

Table 6 — Annual human share of content edits regressed on year, 2010–2025, sixteen annual observations. 2025 covers January to August only.

Edition	Slope, pp per year	t	R ²	Human share of edits, 2010	Human share of edits, 2025
Kazakh (kk)	+4.80	4.21	0.56	17.2%	95.1%
Uzbek (uz)	+4.45	2.94	0.38	2.3%	75.2%
Kyrgyz (ky)	+2.03	1.81	0.19	29.8%	86.9%
Azerbaijani (az)	+1.86	2.00	0.22	55.9%	83.6%
Tatar (tt)	−1.55	−1.34	0.11	26.3%	17.8%
Russian (ru)	−0.11	−0.30	0.01	60.1%	46.0%

On edits the Kazakh trend is the strongest in the set: +4.80 percentage points a year, $t = 4.21$, and a rise from 17.2 per cent in 2010 to 95.1 per cent in the first eight months of 2025. Taken alone this reads as a decisively human edition.

The annual series says why that reading is too generous. Kazakh bot edits fell from 278,875 in 2011 and 436,448 in 2012 to 2,980 in 2025. Kazakh human edits went 5,422 in 2010, 204,444 in 2011, 80,828 in 2016, 102,969 in 2024, and 57,687 in the eight months of 2025. The human share rose overwhelmingly because the denominator collapsed. Human editing has not grown since the middle of the decade; the bots left.

The same point holds for creation, in Kazakh's favour. No flagged bot has created a Kazakh content page since 2012: the 120,333 bot-created pages are 49,468 from 2011 and

70,865 from 2012 and nothing after. Every Kazakh article created in the last thirteen years was created from a human-flagged account, and no month since 2011 trips the burst rule. Whatever the stock is made of, the recent flow is human.

The size of the community that produces that flow is the sobering part. The mean monthly count of Kazakh accounts making 25 or more content edits was 4.0 in 2010, 44.2 in the campaign year of 2011, fell to 12.8 by 2017, and recovered to 37.1 in 2024 and 36.0 in the first eight months of 2025. We fitted both a linear and a quadratic trend to these sixteen points and neither is distinguishable from noise ($t = 1.45$ and $t = 1.61$ respectively), so we report no trend and state the levels instead. The level is the finding: fourteen years after the campaign, the core Kazakh editing community has not returned to the size it reached during it.

The following is the author's own account, written before seeing the results of this study (Telegram, 20 and 23 September 2026), lightly edited and translated.

I open the English Wikipedia for almost everything. It has a large community that edits often, so it is usually the most current, and its articles are the most complete. My schoolwork is in English too, so if I need to quote something it is easiest to take it from there. My classmates do the same, except the ones whose English is weak; they use the Russian Wikipedia. The Kazakh Wikipedia we open very rarely, because there is not much in it. The one exception is when we need something from the history of Kazakhstan, because for that the other languages have little. Comparing three articles myself showed the same thing from the other side. On Astana the Russian article had 19 sections against 14 in Kazakh and was more recent, and the Kazakh one looked translated. On Mathematics the Russian article had 15 sections against 5, but the Kazakh one read as if it had been written in Kazakh rather than translated. On Gennady Golovkin the Russian article had 16 sections against 8, and the Kazakh article was again original rather than translated, with more attention to his state awards where the Russian one gave more space to international ones.

The following is the author's reaction after reading the results (Telegram, 25 September 2026), lightly edited and translated.

What surprised me most is that since 2013 not a single article has been created by a bot. I also did not expect that the rise in the human share of editing is not explained by more people arriving, but by bots leaving. And the most interesting thing is that the number of editors has actually fallen compared with 2011.

6. Limitations

The editor classes are properties of accounts. An unflagged account running a script is counted as human until the burst rule catches it, and a flagged bot performing hand-guided work is counted as a bot throughout. The burst rule narrows the first error and does nothing about the second.

The burst excess is an upper bound on unflagged automation, not an estimate of it. A genuine coordinated human effort — a school assignment, an edit-a-thon, a translation drive — produces the same shape in creation counts, and the behavioural corroborators reduce but do not eliminate that ambiguity: a hurried human stub is also short and also rarely revisited. The honest reading of Kazakh is a range, 27.2 to 48.3 per cent human creation, with our evidence favouring the lower half.

Steady automation is invisible to the rule by construction. An account creating two hundred pages a month for a decade never exceeds its own trailing median and is never flagged. The Uzbek and Russian placebo results (45 and 47 per cent) indicate that some flagged bot work in those editions has exactly this shape, so our corrections there are conservative.

The window begins in January 2010 and covers 70 to 99 per cent of each edition's live stock. The Tatar figure of 70 per cent is the weakest, so Tatar's position in Table 5 is the least secure in the set.

We measure creation and edits, not quality. A human-written page can be worse than a well-generated one, and nothing here should be read as a claim about the content of any article. The quality question would require article-level assessment, which this design does not attempt.

Finally, we do not identify any account, campaign or organisation. The attribution of the 2011 Kazakh bursts to the documented WikiBilim campaign rests on the coincidence of dates with a published record [12], not on account-level evidence, and should be treated as consistent with the data rather than established by it.

7. Conclusion

Of the content pages the Kazakh Wikipedia gained between 2010 and 2025, flagged bots created 50.4 per cent. Reading the burst signature as automation, the share created by a human being falls to 27.2 per cent, and that estimate holds across a tenfold range of the detection threshold. The same exercise puts the Kyrgyz figure at 16.6 per cent and the Tatar figure at 9.7 per cent, while the Azerbaijani edition, of near-identical size to the Kazakh one, was built almost entirely by hand.

The methodological result is the transferable one. The platform's own bot flag is a lower bound on automation and a poor one in exactly the editions where the question matters most, because mass creation also runs through ordinary accounts. The correction is cheap: monthly creation counts by editor class, a trailing-median rule, and two behavioural cross-checks that are already in the same API. It is falsifiable, it stays silent on the editions where nothing happened, and it recovers 87 to 97 per cent of known bot creation when used as a placebo.

For the specific argument this paper started from, the finding cuts both ways and we report both directions. Roughly half to three quarters of the Kazakh article counter is not human work, so the counter should not be quoted as a measure of Kazakh-language knowledge production without this decomposition [1]. But no bot has created a Kazakh article since 2012, the human share of editing has risen every few years since, and the recent flow is entirely human. What has not grown is the number of people doing it: the core Kazakh editing community is smaller today than during the 2011 campaign. The honest summary is that Kazakh Wikipedia's problem is no longer its bots. It is that roughly three dozen people a month are maintaining a quarter of a million articles, half of which somebody else's program wrote fifteen years ago.

Data availability and replication

All data are public and require no account. The pull script (`pull.py`), the analysis script (`analysis.py`) and the complete machine-readable results (`results.json`) accompany this manuscript. Running `pull.py` then `analysis.py` reproduces every number in every table from the Wikimedia Analytics API. The raw API payloads as retrieved on 23 September 2026 are included so that the tables can be reproduced without network access, and so that any later revision of the underlying data lake is detectable by comparison.

Declarations

Funding: none. Conflicts of interest: none. Ethics: the study uses only aggregate public counts; no personal data, no individual accounts and no human subjects are involved. Use of AI tools: the author used AI assistance for data retrieval code, statistical computation and language editing; the research question, the interpretation of results and the conclusions are the author's own, and every reported number traces to the saved scripts and payloads above.

References:

1. Zharkynbekova S., Ayupova G., Galiyeva B., Shakhputova Z., Zabrodskaia A. Digital linguistic sovereignty and Kazakh language policy in the digital environment // *Frontiers in Psychology*. 2026. DOI: 10.3389/fpsyg.2026.1858650.
2. Kornai A. Digital Language Death // *PLoS ONE*. 2013. Vol. 8, No. 10. e77056. DOI: 10.1371/journal.pone.0077056.
3. Steiner T. Bots vs. Wikipedians, Anons vs. Logged-Ins // *Proceedings of the 23rd International Conference on World Wide Web Companion (WWW '14 Companion)*. 2014. arXiv:1402.0412.
4. Steiner T. Bots vs. Wikipedians, Anons vs. Logged-Ins (Redux): A Global Study of Edit Activity on Wikipedia and Wikidata // *Proceedings of the International Symposium on Open Collaboration (OpenSym '14)*. 2014. DOI: 10.1145/2641580.2641613.
5. Tsvetkova M., García-Gavilanes R., Floridi L., Yasseri T. Even good bots fight: The case of Wikipedia // *PLoS ONE*. 2017. Vol. 12, No. 2. e0171774. DOI: 10.1371/journal.pone.0171774.
6. Halfaker A., Riedl J. Bots and Cyborgs: Wikipedia's Immune System // *Computer*. 2012. Vol. 45, No. 3. P. 79–82.
7. Geiger R. S. The Lives of Bots // Lovink G., Tkacz N. (eds.) *Critical Point of View: A Wikipedia Reader*. Amsterdam: Institute of Network Cultures, 2011.
8. Zheng L., Albano C. M., Vora N. M., Mai F., Nickerson J. V. The Roles Bots Play in Wikipedia // *Proceedings of the ACM on Human-Computer Interaction*. 2019. Vol. 3, CSCW, Article 215. DOI: 10.1145/3359317.
9. Reeves N., Simperl E. Machines in the Margins: A Systematic Review of Automated Content Generation for Wikipedia. 2025. arXiv:2509.22443.
10. Wikimedia Foundation. Analytics Query Service (Wikistats 2) REST API. URL: https://wikimedia.org/api/rest_v1/ (accessed 23.09.2026).
11. Wikipedia's largest non-English version was created by a bot. Generative AI poses new problems // *ABC News*. 15.04.2025. URL: <https://www.abc.net.au/news/science/2025-04-15/wikipedia-cebuano-lsjbot-ai-article-generation-non-english/105123090> (accessed 23.09.2026).
12. Davis L. Class assignment inspired Wikipedian of the Year to grow Kazakh Wikipedia // *Wikimedia Foundation, Diff*. 05.07.2012. URL: <https://diff.wikimedia.org/2012/07/05/class-assignment-inspired-wikipedian-of-the-year-to-grow-kazakh-wikipedia/> (accessed 23.09.2026).

МАЗМҰНЫ – СОДЕРЖАНИЕ – CONTENT

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

Төлегенов Тимур Сейфуллаулы

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

Alexander Gorevoy

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

Rakhimov Djalil

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

Kazhgerey Asan

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

Tagimova Alua

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

Wei Wei, Mynzhanova Gulzhakan, Niyetalina Gaukhar

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

Xu Fumin, Wei Wei, Ruziyeva Elvira, Moldabekova A.

**FROM CULTURAL RESOURCES TO GLOBAL CONSUMER VALUE: THE BRIDGE–IMC MODEL
AND THE INTERNATIONALIZATION OF BLACK MYTH: WUKONG UNDER THE BELT AND
ROAD INITIATIVE 64**

Ibraimkhan Alikhan

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

Yerboluly Madi

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

Салимбаев Ербол Аманжолович

**АНАЛИЗ ИНВЕСТИЦИОННОЙ АКТИВНОСТИ КАК ИНДИКАТОР ИЗМЕНЕНИЙ
ПОТЕНЦИАЛЬНОГО СПРОСА НА КОНСАЛТИНГОВЫЕ УСЛУГИ 99**

Rambergenov Ramil

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

Abdesh Dinara

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

Amangeldi Aidana

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

Cheng Che	
BUILD, BUY, OR PARTNER? ENTRY MODE SELECTION AND RISK GOVERNANCE FOR WIND POWER INVESTMENT IN KAZAKHSTAN	133
Wang Liang	
FROM CAPACITY EXPANSION TO SYSTEM VALUE: COORDINATED DEVELOPMENT OF RENEWABLE ENERGY, GRID FLEXIBILITY AND ENERGY STORAGE IN KAZAKHSTAN	146
Chen Jun	
CONSTRUCTION-PERIOD CREDIT RISK MANAGEMENT FOR LARGE INFRASTRUCTURE PROJECT FINANCE IN KAZAKHSTAN	159
Alniyaz Danial	
HOW MUCH FUEL DOES A HYBRID REALLY SAVE? A SAME-MODEL COMPARISON OF HYBRID AND GASOLINE CARS IN US FUEL ECONOMY	173

ТЕХНИКАЛЫҚ ҒЫЛЫМДАР - ТЕХНИЧЕСКИЕ НАУКИ - TECHNICAL SCIENCE

Nurmakhan Almira	
ADVANTAGES OF FIBER OPTIC COMMUNICATION OVER TRADITIONAL COPPER CABLES	180
Kozhabai Bereke	
SECURITY CHALLENGES IN IOT-BASED TELECOMMUNICATION NETWORKS	188
Azamatuly Yerassyl	
WHY AI-GENERATED CODE FAILS A BENCHMARK: A FAILURE ANALYSIS OF THREE OPEN LANGUAGE MODELS ON 30 HUMANEVAL+ TASKS	197
Nurzhan Sanzhar	
A FEW SHELLS, A FEW OPERATORS: WHERE LOW EARTH ORBIT IS CROWDED IN AN ALTITUDE-INCLINATION MAP OF 16,620 ACTIVE SATELLITES	206
Askarbek Nurislam	
CAN THE WEATHER TELL WHICH DAYS IN ASTANA WILL BE POLLUTED? CLASSIFYING PM2.5 EXCEEDANCE DAYS WITH MACHINE LEARNING AND CHECKING THE DATA AGAINST A GROUND MONITOR	212
Iliyas Arman	
SHOOTING ON THE MOVE: HOW MUCH AIM A ROBOT LOSES BY IGNORING ITS OWN VELOCITY, AND WHETHER A VIRTUAL-TARGET CORRECTION RECOVERS IT	223
Ануар Әлімжан Бахытжанұлы, Сексенбаева Айша Курманбековна	
СРАВНИТЕЛЬНЫЙ АНАЛИЗ СТАНДАРТОВ И МЕТОДИК ОБУЧЕНИЯ КИБЕРБЕЗОПАСНОСТИ: NICE FRAMEWORK, ECSF И ПРАКТИКО- ОРИЕНТИРОВАННЫЕ ТЕХНОЛОГИИ	229
Mukhamedzhan Yerganat	
GROWTH OF GITHUB DEVELOPER COMMUNITIES IN KAZAKHSTAN, UZBEKISTAN AND KYRGYZSTAN, 2020-2026: EVIDENCE FROM THE GITHUB INNOVATION GRAPH	239

ПӘНДЕР АРАЛЫҚ ҒЫЛЫМДАР – МЕЖДИСЦИПЛИНАРНЫЕ НАУКИ – INTERDISCIPLINARY SCIENCES

Albiyev Asser	
HOW MUCH OF AN ARTICLE COUNT IS HUMAN WORK? DECOMPOSING THE CREATION OF FIVE WIKIPEDIAS, 2010–2025	251
Amangeldy Nurbolat	
HOW MUCH OF KAZAKHSTAN'S URBAN BUILT STOCK PREDATES ITS SEISMIC CODES, AND WHERE DOES THAT STOCK SIT RELATIVE TO HAZARD?	261

Torekhan Sabrina

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

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

Мұқатаева Айзат Амантайқызы, Стамбекова Жанат Байғазықызы, Слямханова Гүлім Болатқызы, Қызырбек Аяулым Ерболқызы

ЖАСАНДЫ ИНТЕЛЛЕКТ ДӘУІРІНДЕГІ ҒЫЛЫМИ ТАНЫМ: ПОППЕР, КУН ЖӘНЕ ЛАКАТОС ИДЕЯЛАРЫ ТҮРҒЫСЫНАН ТАЛДАУ	272
--	------------

Мұқатаева Айзат Амантайқызы, Сапарханова Мадина Төлегенқызы, Окенова Аружан Муратхановна, Тұрсынғазыева Ақерке Ерболатқызы

ШӘКӘРІМ ФИЛОСОФИЯСЫНДАҒЫ ЕРКІНДІК ЖӘНЕ МОРАЛЬДЫҚ ТАҢДАУ МӘСЕЛЕСІ	282
---	------------

Komekova Saida

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

Zhumagali Yerbol

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

Abdrakhmanova Madina

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

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

Kaldarbek Mergen

FAMILY OR SCHOOL? SOCIOECONOMIC COMPOSITION AND MATHEMATICS ACHIEVEMENT IN KAZAKHSTAN, PISA 2022 AND 2025	316
--	------------

Zhenis Adilzhan

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

Базарбаева Айнура Бектуровна

ДЕТСКИЙ ФОЛЬКЛОР И ПОДВИЖНЫЕ ИГРЫ КАК ИНСТРУМЕНТЫ ФОРМИРОВАНИЯ ОСНОВ ПРАВОВОЙ КУЛЬТУРЫ У ДЕТЕЙ СТАРШЕГО ДОШКОЛЬНОГО ВОЗРАСТА	330
---	------------

БИОЛОГИЯ ҒЫЛЫМДАР - БИОЛОГИЧЕСКИЕ НАУКИ - BIOLOGICAL SCIENCES

Bazhenova Diana

GROWTH DYNAMICS OF STERLET (<i>Acipenser ruthenus</i> L.) REARED IN A RECIRCULATING AQUACULTURE SYSTEM	333
---	------------

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

Төлеп Жазира Маратқызы

ГАСТРОНОМИЯЛЫҚ ФЕСТИВАЛЬДЕРДІҢ ТУРИСТІК ДЕСТИНАЦИЯ ТАРТЫМДЫЛЫҒЫН ҚАЛЫПТАСТЫРУДАҒЫ РӨЛІ МЕН ДАМУ ҮРДІСТЕРІ	340
--	------------

ЕСТЕСТВЕННЫЕ НАУКИ - ЖАРАТЫЛЫСТАНУ ҒЫЛЫМДАРЫ - NATURAL SCIENCES

Нургельдина Асель Ермаковна, Серикбаева Жанерке Алимжановна

ШЕКАРАСЫЗ ФИЗИКА: ӘР ЕЛДЕГІ ФИЗИКАЛЫҚ БІЛІМДІ САЛЫСТЫРУ 347