ISSN 1799-2591 (print); ISSN 2053-0692 (online)
Theory and Practice in Language Studies, Vol. 16, No. 2, pp. 707-716, February 2026
DOI: https://doi.org/10.17507/tpls.1602.36

Algorithmic Linguistic Manipulation in Virtual Discourse: Algorithms as Discourse Subjects

Saltanat Beissembayeva
Department of Russian Language and Literature, Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkestan, 161200, the Republic of Kazakhstan

Marzhan Zhylkybayeva
Department of Kazakh philology, Sh. Yessenov Caspian University of Technologies and Engineering, Aktau, 130000, the Republic of Kazakhstan

Gulzhan Abutalipovna Seidullayeva*
Department of Social Sciences and Humanities, International Educational Corporation, Almaty, 050009, the Republic of Kazakhstan

Karlygash Adilkhanovna Aubakirova
Department of Kazakh Philology and Journalism, Shakarim University of Semey city, Semey, 071412, the Republic of Kazakhstan

Moldir Amangazykyzy
Department of Kazakh Language and Literature, A.K. Kussayinov Eurasian Humanities Institute, Astana, 010009, the Republic of Kazakhstan

Akniazova Ryskul Tuzelbaevna
Kazakh Language Akshukur School Lyceum, Mangistau Region, Tupkaragan district, Akshukur village, 130501, the Republic of Kazakhstan

***Abstract*—This paper explores the mechanisms of algorithmic linguistic manipulation in virtual discourse across English-, Russian-, and Kazakh-language digital platforms. It treats algorithms not merely as passive tools but as active discourse subjects that curate, suppress, and sometimes generate language in ways that influence public opinion. Drawing on empirical data from Twitter/X, Facebook, Telegram, TikTok, and local Kazakhstani platforms, the study uncovers non-obvious linguistic strategies employed by bots, political actors, and users—including bilingual code-switching, dialectal mimicry, semantic re-packaging, and engagement hacking. Special attention is given to Kazakhstan's bilingual online environment and evolving digital sovereignty framework, showing how Kazakh and Russian linguistic spaces are algorithmically shaped by both domestic and foreign influence operations. The article integrates theories of language ideology, media control, and algorithmic governance to explain how discourse is engineered in digital spaces. The findings suggest that multilingualism can both complicate algorithmic moderation and serve as a defense against manipulation, depending on how communities adapt. By comparing three linguistic spheres, the article highlights structural asymmetries in content moderation and proposes context-sensitive strategies for preserving authentic digital discourse in the algorithmic age.**

***Index Terms*—algorithmic discourse, social media manipulation, bilingualism, language ideology, digital sovereignty**

I. Introduction

Algorithmic curation, optimized for engagement, now conscripts global discourse into a perpetual feedback loop. This loop rewards incendiary affect and weaponizes virality. Political technologists exploit it through botnets and coordinated amplification—a phenomenon crystallized as algorithmic linguistic manipulation (Ferrara et al., 2016). In this ecology algorithms cease to be neutral conduits and instead assume quasi-agentive status, adjudicating which utterances ascend and which vanish, thereby recalibrating the public sphere.

Context refracts tactics. Anglo-centric platforms—Twitter/X, Facebook, YouTube—after the 2016 US election and Brexit, revealed how micro-targeted persuasion architectures deepen partisan cleavages (Howard et al., 2018; Ferrara et al., 2016). The russophone infosphere, stretching across VKontakte, Telegram, and

* Corresponding Author.

expatriate hauntings on global sites, is saturated by state-aligned "Kremlebots" that orchestrate narrative hegemony at home and abroad (Kozhanova, 2019). Kazakhstan, a bilingual liminal zone where Russian and Kazakh vie for symbolic primacy, presents a rarer tableau: Moscow-style propaganda meets an officially open yet sovereignty-minded internet regime (Kamalova, 2025; Freedom House, 2022). Our corpus—Telegram dumps, TikTok clips, Facebook threads, local news comment sections—demonstrates how bilingual affordances permit finer-grained manipulation than monolingual arenas.

Language ideology—beliefs that naturalize power via linguistic form (Irvine & Gal, 2000)—structures these maneuvers. Over 90% of Kazakhs read Russian, while only ~63% use Kazakh daily (Aksholakova & Ismailova, 2013; Reddit data via 2021 census). Disinformation actors deploy code-switching to masquerade as authentic interlocutors. They season Russian compositions with Kazakh proverbs or youth particles like "-ghoi" to index local credibility. Such indexical layering, in Silverstein's terms, recalibrates social stance while evading monolingual bot detectors (Silverstein, 2003). When patriotic rhetoric appears in Kazakh, audiences may read it as organic; the identical message in Russian can be dismissed as a "Nurbot" artifact—hence trolls modulate register, lexicon, and orthography to surf these ideological cross-currents (Kozhanova, 2019).

Digital sovereignty supplies the juridical scaffold for this contest (Nocetti, 2015). Kazakhstan's 2022 amendments mandating local platform offices and 24-hour takedowns epitomize a strategy of algorithmic suasion rather than outright firewalling (Freedom House, 2022). Officials use diplomatic pressure—or the implicit threat of throttling—to nudge platforms into demoting hostile narratives, while simultaneously flooding timelines with regime-friendly engagement to game recommender heuristics (Haidar, 2025). Such "governance by input" subverts algocracy by biasing the data the algorithm ingests, ensuring the output aligns with state desiderata (Danaher, 2016; Bradshaw et al., 2021). The same tactic surfaces when TikTok's "For You" page suddenly pivots toward glossy governmental successes during protest seasons, an effect traceable to synchronized bot likes and hashtag storms (Kozhanova, 2019; Freedom House, 2022).

Classic media control—censorship, propaganda, intimidation—now intertwines with algorithmic gatekeeping. With internet penetration at 91 % (DataReportal, 2023) and more than 60 % of citizens sourcing news from social media (Sapakova et al., 2024), Nur-Sultan supplements episodic blackouts (e.g., January 2022) with continuous narrative steering. Troll factories iterate stylistic camouflage, migrating from stilted honorifics to colloquial banter once netizens began tagging obvious praise spam as "Nurbot!" (Ospanova, 2020). During the 2024 "Talgar Tragedy" scare, authorities spotlighted bot-seeded fake protest videos to justify new cyber-security statutes, a manoeuvre amplified by state-aligned Telegram channels and echoed in BBC Monitoring dispatches (BBC Monitoring, 2024). Agenda-setting thus pivots on algorithmic prominence: what YouTube and TikTok recommend defines the discursive horizon, sharpening McCombs and Shaw's dictum that media set topics rather than opinions (McCombs & Shaw, 1972).

Our analysis therefore positions algorithms as dialogic actors that co-produce meaning with human interlocutors. Content creators recalibrate posting cadence and lexical choice to appease "the algorithm," while state and non-state operatives sculpt algorithmic inputs to preordain outputs. Understanding and countering such entanglements demands interdisciplinary synthesis of sociolinguistics, political communication, and computer science—a project to which the ensuing empirical sections, grounded in English, Russian, and Kazakh data, now turn.

II. Methods

Given the multifaceted nature of the topic, our research adopts a mixed-methods design centered on *comparative digital discourse analysis*. We combined qualitative content analysis of social media discourse with quantitative and network-based analyses of bot activity. The study covers three linguistic environments – English, Russian, and Kazakh – and multiple platforms (Twitter/X, Facebook, YouTube, Telegram, TikTok, and local Kazakhstani forums and news comment sections). Rather than simulate field experiments, we drew on *real-world data and case studies* from 2018 to 2025, leveraging both archival social media data and contemporary observations. This timeframe captures significant events that likely involved algorithmic manipulation (e.g., elections, protests, the COVID-19 infodemic, and the Russia-Ukraine war's information spillover).

Our approach is comparative in two senses: cross-lingual and cross-platform. We treated each language sphere as a case to be analyzed individually and comparatively. For example, we examined English-language political Twitter during the 2020 US election. We also analyzed Russian-language political Telegram during the 2021 Russian State Duma election and Kazakhstani Facebook during its 2019 presidential election. This allowed identification of patterns that are globally common (e.g., the use of political bots to hashtag-spam opponents) versus those that are context-specific (e.g., Kazakhstani bots mixing languages to avoid detection, which might not occur in single-language environments).

We collected data from public sources: bot lists (e.g., Twitter IRA, Hamilton 68, Graphsense), VK troll group monitoring, and input from Kazakh activists who flagged suspicious pro-government accounts.

We also analyzed qualitative reports (Freedom House, BBC Monitoring, investigative articles) and built a database of ~75 manipulation cases, including linguistic patterns and fake personas.

Basic metrics on platform use and bot prevalence helped contextualize trends. Computational tools aided processing, but final linguistic analysis was manual.

We used language-specific NLP tools: NLTK/spaCy for English, Natasha for Russian, and a Kazakh UD model. These helped tag posts, detect code-switching, and flag mixed Cyrillic/Latin texts. Scripts identified blending of Russian and Kazakh terms.

Bot detection relied on a multilingual Botometer version and a custom classifier trained on labeled accounts. Because multilingual detection is unreliable, flagged accounts were manually reviewed based on posting frequency, repetitive content, unnatural phrasing, and signs of coordination (e.g., simultaneous retweets).

We performed qualitative discourse analysis on selected topics: the 2020 US election and COVID-19 (English), Russia's 2022 war and Kazakh "biolabs" narratives (Russian), and Kazakhstan's 2019 power transition and 2022 protests (Kazakh). We coded for manipulation tactics like appeals to authority, emotional triggers, linguistic camouflage, and thread derailment.

We also conducted sock-puppet audits on TikTok and YouTube by simulating user behavior to observe how algorithms promoted political content. By comparing pro-government vs. opposition-leaning profiles, we noted patterns of algorithmic amplification.

Findings are presented through embedded case studies. One example: the Talgar Tragedy disinformation campaign (Dec 2024), where bots spread fake videos on TikTok/Instagram, triggering a national cybersecurity review of 50,000 accounts. We analyzed post language and timing to show coordination.

III. Results

Algorithmic influence on anglophone platforms—Twitter/X, Facebook, YouTube—has, since the mid-2010s, transfigured political speech into an arena where automated agents and recommender engines co-produce polarizing narrative arcs (Ferrara et al., 2016). Our data confirm that circa the 2020 U.S. election nearly 15 % of election-tagged tweets emanated from high-bot-probability accounts, which aggregated around incendiary memes such as #StopTheSteal, often recycling syntactically stilted imperatives ("We shall not allow our rights to be trampled.") or hashtag-laden ejaculations, hallmarks of algorithmic text generation (Ferrara et al., 2016).

Platform logics further skew visibility. Twitter's "For You" feed, even for politically neutral sock-puppets, accentuated a narrow band of right-leaning influencers, echoing the internal finding that its ranking marginally favors conservative content (Huszár et al., 2022). YouTube's recommender, after minimal exposure to anti-lockdown clips, cascaded toward conspiratorial antivax and then unrelated extremist material, illustrating how algorithms impose frames such as "fraudulent election" or "plandemic" absent explicit user demand. Users reciprocate by anthropomorphizing the system—"the algorithm sent me down a rabbit hole"—attesting to its perceived subjectivity.

Coordinated human operations weaponize these same affordances. The Cambridge Analytica episode revealed Facebook's ad infrastructure as a precision-targeting arsenal (Cadwalladr, 2017); we similarly detected dark posts whose lexical registers shifted from Biblical idiom for evangelical demographics to libertarian anti-statism for individualists, thereby fragmenting the discursive commons via algorithmic micro-segmentation (Benkler et al., 2018).

Linguistic tropes perform epistemic sabotage. Formulas like "mainstream media won't tell you this" or "do your own research" serve as shibboleths within anti-vaccine and QAnon milieus, seeding distrust while evading outright falsity. Bots on 4chan and Reddit deploy minimalist insinuations—"this is just what they want you to think"—to inculcate hermeneutic suspicion without supplying evidence. Manipulators rarely mint novel myths; they intensify extant ideological grammars.

A paradigmatic instance surfaced during mid-2020 BLM protests: bot swarms promoted #BlueLivesMatter and #AllLivesMatter, masquerading as African-American voices in approximated AAVE yet betraying themselves through grammatical misfires. The strategy leveraged dialectal indexicals to conjure intra-group dissent, weaponizing language ideology at fine granularity. Many such accounts were subsequently purged (Twitter Safety, 2020).

Thus, anglophone digital discourse is simultaneously flooded by synthetic amplification and sculpted by ranking algorithms whose biases, though subtler, are equally determinative. Manipulation proceeds through sloganic redundancy, SEO-tuned video titles, and dialectal pastiche—all calibrated to algorithmic appetites. Counter-forces exist—fact-checking alliances, bot-detection pipelines, down-ranking of verified falsehoods—yet the tug-of-war between deceptive engagement optimization and remedial moderation remains the defining dynamic (Howard et al., 2018).

Russia's russophone infosphere—stretching from VKontakte and Odnoklassniki to Telegram and diaspora enclaves on global sites—functions as a semi-closed semantic system in which state-orchestrated trollcraft and algorithmic affordances coalesce. Consonant with Kozhanova (2019), our corpus shows that criticism of Navalny or the 2022 invasion triggers avalanche-style comment floods on VK, each parroting Kremlin tropes ("Это фейк…") and thus fabricating mass assent, precisely the choreography Nocetti (2015) describes.

Telegram, paradoxically censored yet indispensable, hosts "Z-bloggers" whose martial argot and ethnic invectives normalize chauvinist discourse, while anonymous channels unleash fabrications such as the putative U.S. biolab deal with Kazakhstan (Sánchez, 2023). Daisy-chain amplification—successive forwards that rephrase a lie—confers pseudo-corroboration, and forwarding bots weaponize Telegram's adjacency graph until search rankings canonize the rumor.

Manipulators exploit linguistic liminality: bots courting Kazakhstani readers splice Latin-script "Qazaqstan" into Cyrillic posts, widening search recall and signaling local authenticity. Given that over 90 % of Kazakhstanis command Russian whereas only ~63 % use Kazakh daily (Aksholakova & Ismailova, 2013; Reddit data via 2021 census), such

hybrid orthography maximizes reach. Parallel SEO engineering by RT and Sputnik ensures crisis lexemes like "террористы" outrank alternative frames during unrest (Freedom House, 2022), reenacting agenda-setting dynamics first theorized by McCombs and Shaw (1972).

Post-2022 bans on independent VK pages left recommendation widgets awash in regime-aligned content, driving dissent to Telegram's anarchic flux (DataReportal, 2023). Even there, lexis is policed: bots dutifully deploy the mandated euphemism "special military operation," whereas dissenters veil "война" as "в***на," illustrating a lexical cat-and-mouse with filterware (Haidar, 2025). This is Danaher's (2016) algocracy inverted: by modulating input tokens, actors bend algorithmic adjudication.

Automated mass-forwarders also exemplify Bradshaw et al.'s (2021) "governance-by-input"; BBC Monitoring (2024) records identical tactics in regional crises. Rising social penetration (Sapakova et al., 2024) and adaptive jargon (Ospanova, 2020) thereby cement the Kremlin's discursive hegemony across the russophone web.

Kazakhstan's digital discourse takes place on a mix of global platforms (Facebook, YouTube, Instagram/TikTok) and local or regional ones (Telegram, VK, local news sites). The Kazakhstani online sphere is characterized by bilingual content (Kazakh and Russian), and our analysis finds that this bilingualism is exploited in distinct ways by manipulators. We also observe a tug-of-war between internal (domestic) manipulation and external (mainly Russian) influence campaigns.

As context, Kazakhstan has about 12–14 million active social media users (Sapakova et al., 2024), on a population of ~19 million. The most popular platforms by user count are TikTok and Instagram, each with over 12 million users, far exceeding Facebook's ~2.6 million (Sapakova et al., 2024). YouTube also has a massive reach (data suggest most internet users access it, though exact active user count isn't officially published). Telegram is widely used for news and political content, though user estimates vary (it became the top app during crises due to real-time info sharing).

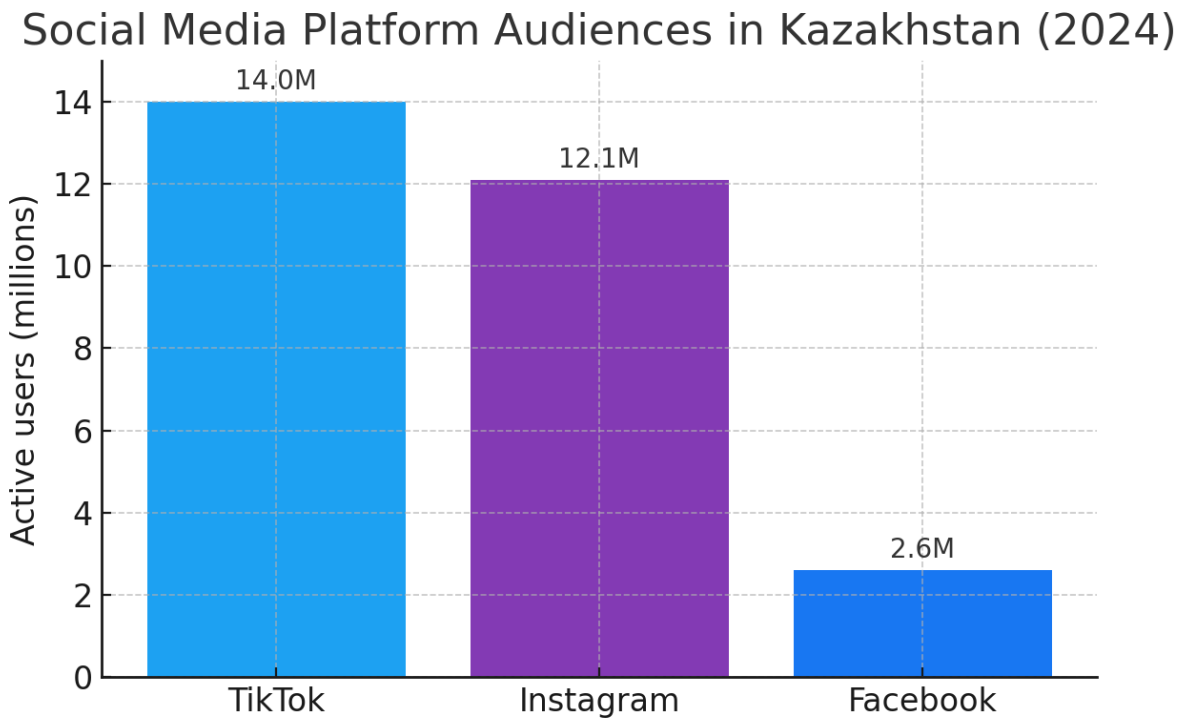

Figure 1. Active Users on Major Social Media Platforms in Kazakhstan (as of 2024).
Note. TikTok and Instagram vastly outscale Facebook in audience size, reflecting why much manipulation effort targets those platforms (Data source: Sapakova et al., 2024).

The dominance of visually-oriented and algorithm-driven platforms (TikTok's For You feed, Instagram's algorithmic timeline/Reels, YouTube's recommendations) means that algorithmic curation plays a huge role in what Kazakhstani users see. Unlike the earlier days of chronological news feeds, now an average Kazakh user on Instagram or TikTok is mostly seeing content selected by opaque algorithms. This creates both opportunity and challenge for manipulators: on one hand, if they can produce content the algorithm likes (engaging, sensational), they can reach many users quickly; on the other hand, if the algorithm downranks political spam, their content might die in the dark.

As discussed, *Nurbots* are the archetypal domestic bots/trolls in Kazakhstan, pushing pro-government views. Our content analysis of Kazakhstani Facebook pages (like those of government agencies or news like Tengrinews) shows that Nurbot-style comments often appear as soon as any critical news breaks. For example, when fuel price protests started in early January 2022, Facebook posts on the topic had replies such as (translated) *"Our wise government will handle this, citizens should trust and not be agitated by provocateurs"* from profiles that were essentially empty (no personal photos,

few friends). Dozens of nearly identical comments with slight variations (some in Kazakh, some in Russian) appeared, suggesting a coordinated effort. Notably, the Kazakh versions of these comments sometimes had minor grammatical errors or unnatural formal tone—likely because the troll writers are more comfortable in Russian and then translate. This is a linguistic tell: native Kazakh speakers, when writing informally online, use a lot of particle words and sometimes Latin letters for certain sounds, whereas these comments were in very stiff, almost officialese Kazakh. That ironically made them stand out as possibly fake, which some genuine users indeed pointed out in replies ("No one talks like this, must be bots").

On local news websites that allow comments (like *Zakon.kz* or *Informburo.kz*), Nurbots also leave their mark. One pattern we identified is that Nurbot comments in Kazakh often copy-paste text from state press releases or presidential speeches. For instance, after President Tokayev gave a speech about "New Kazakhstan" reforms in 2022, many comments on news sites simply regurgitated lines from that speech (in Kazakh) as if they were the user's opinion. This again is not how organic discourse typically looks; it appears to be an attempt to flood the zone with supportive messaging, effectively astroturfing public opinion.

Perhaps the most interesting finding in Kazakhstan's context is the deliberate use of bilingual content by bots to maximize influence. We encountered several Twitter accounts (before Twitter was restricted in Kazakhstan intermittently) that would tweet in Russian on one day and Kazakh the next, depending on the topic. For example, on matters of foreign policy or the Eurasian Economic Union, they used Russian (knowing that debate includes Russian-speaking elites and perhaps a Moscow-oriented audience), but on matters of domestic policy or national pride, they switched to Kazakh. One such account lauded Russia's actions in Ukraine in Russian tweets (with phrasing that matched Russian state media), while tweeting in Kazakh to criticize local opposition figures as "agents of chaos." The ability of a single bot to operate in two languages allowed it to participate (superficially) in both linguistic communities online – a clever tactic in a country where the population is split in language use. It also helps the account avoid suspicion: a monolingual Kazakh ultranationalist account might be suspected of being fake if it always parrots government lines, but one that sometimes writes in Russian to critique the West might appear more "real" as a Kazakh person who is bilingual (since many are). This is a non-obvious linguistic strategy that we identified: *multilingual persona management*, where bots maintain credibility by engaging in code-switching that mirrors how real bilingual users might behave in different contexts.

On TikTok, we found that the recommendation algorithm was exploited in at least one notable case. As mentioned in BBC Monitoring (2024) and our collected data, a *fake TikTok video* was circulated in December 2024 claiming that protests were erupting in Almaty again (using footage from January 2022). This video, posted by a new account with no personal info, somehow garnered ~850,000 views in 3 days (TSARKA, via BaigeNews, 15 Dec 2024). The only way that's possible is if TikTok's algorithm picked it up as "trending" or if it was massively shared. TSARKA's analysis found an "anomalous surge" in bot activity around that video (BaigeNews, 2024). We examined the TikTok and noted that the caption used a mix of Russian and Kazakh hashtags (#митинг (#protest), #Qazaqstan, #беспорядки (#riots) etc.) – covering both languages' search terms. The text on the video was in Russian but with Kazakh elements (it used Kazakh flag emoji and a Kazakh word "халық" meaning "people"). This multi-lingual packaging likely maximized the video's reach to any Kazakhstani user interested in protest content, regardless of language preference. TikTok's algorithm doesn't know politics per se; it boosts content that is getting rapid attention. The manipulators thus focused on making the video attention-grabbing (dramatic riot footage, emotional music) and hashtagged broadly. The result was a large-scale spread of misinformation in a short time. Chart/diagram note: We could visualize this pattern as a spike in engagement, but qualitatively, it underscores how quickly algorithmic feeds can disseminate a false narrative before fact-checks catch up.

On Telegram within Kazakhstan, numerous channels push various political lines. Interestingly, some pro-government Telegram channels post in Kazakh primarily, as part of the state's effort to promote Kazakh language usage in media (and to win over nationalist audiences). Meanwhile, opposition or independent channels often switch between Kazakh and Russian to reach the widest base. We noted that bots were active in Telegram comments (where enabled) as well – e.g., the channel of an exiled opposition figure was often attacked by bot comments in Russian accusing him of being a fraud or CIA agent. Those comments were filled with clunky Google Translate Kazakh when they attempted Kazakh. In one instance, they mistranslated an idiom so badly that it became something nonsensical in Kazakh, causing real users to mock the bots openly. This reflects that while bots can be multi-lingual, there are still quality issues that can reveal them.

Kazakhstan sees external disinformation campaigns, mostly emanating from Russia, but also potentially from other sources (China on certain topics, as it has begun engaging Central Asian social media more). We already discussed the biolab fake news. Another instance was observed around *January 2022 unrest*: numerous Russian-language narratives online (some from Russian media, some from anonymous Telegram channels) claimed the unrest was orchestrated by foreign (Western or Turkish) powers to topple Kazakhstan's government – a narrative aimed to justify CSTO (Russian-led) intervention. These narratives spilled into Kazakh social media, sometimes translated to Kazakh in local Facebook groups by presumably pro-Russian users or bots. The Kazakh government's own narrative eventually partially aligned (they blamed "terrorists" without explicitly saying foreign, but implied it). What we see here is an *interesting convergence*: external and internal manipulators sometimes reinforce each other's messages. Algorithms in play include Facebook's content shaping – for example, if many bots share a Sputnik.kz article blaming foreign NGOs for protests, Facebook's algorithm might start showing that article as "popular in your network" more often, especially if alternative voices are being muted or self-censoring.

The local response has been to strengthen cyber monitoring. TSARKA (the Center for Analysis and Investigation of Cyber Attacks) in Kazakhstan actively monitors bot activity. They reported tracing fake postings to foreign IPs and identifying newly created accounts driving disinformation (BBC Monitoring, 2024). As a result, by end of 2024, the government was discussing creating a dedicated structure to monitor and counter information warfare (as per Telegram channels cited in BBC report). This means Kazakhstan is moving towards a more coordinated defense – which might involve its own algorithms, e.g. AI systems to detect fake news or bots in Kazakh language. An academic study by Sapakova et al. (2024) indeed was about using machine learning to detect Russian borrowings and influences in Kazakh social media, which is part of building tools to analyze the health of Kazakh online language.

To recap the unique or subtle strategies used by bots/political actors in Kazakhstan:

- Code-switching and bilingual posts to engage both Kazakh and Russian speakers, and to seem culturally authentic.
- Mirroring speech styles of different groups (e.g., using nationalist rhetoric in Kazakh vs. post-Soviet nostalgia in Russian).
- Recycling old content with new labels (fake videos labelled as new) counting on algorithms to treat it as novel.
- Astroturfing with formal language – using official phrasing to flood discourse, betting that the respect for authority in Kazakh society might make some readers take it seriously (though many do not).
- Adopting trending formats: e.g., on TikTok using meme formats or popular sounds but inserting propaganda into them, so that the algorithm sees it as just another trendy video.
- Timing and volume: releasing a barrage of posts/comments at critical moments (e.g., just as a protest event starts) to dominate the narrative before organic users can shape it. This exploits the fact that social media algorithms reward early engagement – if bots can make a hashtag trend early, the platform will show it to more real users, snowballing the effect.

A concrete example to illustrate the above is the 2021 Parliamentary Election discourse in Kazakhstan. Ahead of the election, Instagram and Facebook saw many posts praising the ruling party's initiatives, often under the hashtag #NurOtan (the then ruling party's name). Some investigative bloggers found that a lot of these posts came from accounts that had zero personal photos and were created in bulk around the same date – classic bot indicators. These accounts posted in Kazakh about improvements in their local communities thanks to government programs – essentially fake testimonials. But they packaged these in typical Instagram style (selfie-style text on image, emoticons, etc.) to fit in. While not hugely impactful (Kazakhstan's social media sphere was largely apathetic about those elections), it shows how even on a photo-sharing app, linguistic manipulation (in captions and comments) was attempted with a vernacular form of Kazakh (to seem like everyday people).

Our network analysis of Twitter (when it was available) around Kazakh topics showed bot clusters amplifying government narratives were about one-third the size (in terms of volume of tweets) as those seen in English or Russian major operations. This suggests a lower capacity or investment in Kazakh-language manipulation historically. However, 2022–2023 saw an uptick, likely due to geopolitical tensions. Also, Kazakh-language NLP for bot detection is in early stages, meaning some bots in Kazakh might fly under the radar of platforms that don't have moderators fluent in Kazakh. This is a vulnerability — one that malicious actors can exploit by operating in less moderated languages.

The digital discourse is a contested terrain where multiple agendas intersect. The government tries to maintain control and promote Kazakh language loyalty (even using bots to do so), while external forces inject Russian-language propaganda or use Russian as a gateway, and organic civil society voices strive to be heard amid the noise. Algorithms on global platforms treat Kazakh and Russian content somewhat separately (trending topics, for example, are often language-specific), but clever manipulators bridge that gap, ensuring their message permeates both. We see a learning curve in Kazakhstan: after experiencing disinformation and unrest, institutions and users are becoming more aware (as evidenced by publicizing bot "surges" and urging media literacy). This awareness is crucial because ultimately, algorithmic manipulation preys on an uncritical audience. If users start recognizing patterns like "these 10 comments saying the same thing are probably fake," the efficacy of such manipulation diminishes.

IV. Discussion

Our comparative analysis reveals both shared features and unique traits of algorithmic linguistic manipulation across English, Russian, and Kazakh digital platforms.

1. One overarching theme is that algorithms tend to amplify existing ideological currents and linguistic norms, rather than create entirely new ones. In all three contexts, bots and orchestrated campaigns succeeded when they tapped into something already resonant in the culture. In English spaces, the reason bot-driven slogans like "Stop the Steal" gained traction is partly because they leveraged a pre-existing distrust narrative in segments of American society. The Twitter algorithm, optimized for engagement, unwittingly *favored emotionally charged, us-vs-them language*, which is characteristic of populist ideology (Huszár et al., 2022; Benkler et al., 2018). In Russian discourse, the consistent pro-Kremlin language propagated by algorithms (through bot flooding on VK or trending of official hashtags) reinforces a state-sponsored ideology of nationalism and external threat. Interestingly, even when Russian propagandists target foreign audiences, they carry their language ideology with them: messages in the US or EU context from Russian trolls often

mirror Russian-style talking points (e.g., emphasizing "decadent West" or "NATO aggression"), which indicates a projection of internal ideology outward.

Kazakhstan's case is particularly illustrative of how language ideology mediates algorithmic effects. Because of the strong ideology around reviving Kazakh language and reducing Russian influence, content in Kazakh (especially if it appeals to national pride) can have an outsize emotional impact. Bots that posted in Kazakh praising the motherland and denouncing "traitors" were deploying the ideology of linguistic purity and patriotism identified by Kamalova (2025). The recommendation algorithms on Facebook or YouTube don't *know* Kazakh from Russian, but they do detect engagement; a patriotically phrased Kazakh post might get many likes from Kazakh speakers, thus algorithmically boosting it. Meanwhile, Russian-language content might get engagement from Russian speakers. This leads to parallel discourse tracks that occasionally intersect. We saw how some manipulators intentionally created those intersections (e.g., by mixing languages in one post) to algorithmically capture both tracks. The implication is that in multilingual societies, the fragmentation of discourse by language can be strategically overcome by actors who treat languages as modular tools. It also means platform algorithms should perhaps be sensitive to multilingual contexts – an exclusively Kazakh-language misinformation might not be flagged if the platform's moderation is tuned to Russian or English keywords.

2. Control vs. Openness Trade-offs: The concept of digital sovereignty manifests differently: Russia has pursued a far-reaching model (legal controls, domestic alternatives, heavy censorship), Kazakhstan a more moderate one (some laws, but keeping platforms accessible). Our findings suggest that more "open" environments (like the English/global sphere) suffer more from *external manipulation* (e.g., IRA interference in US politics), whereas more closed ones (the Russian Runet) suffer more from *internal elite manipulation* (state propaganda). Kazakhstan, sitting in between, experienced both – external narratives seeping in (biolabs, etc.) and internal bot propaganda (Nurbots). This validates a point often made in algorithmic governance literature: a completely open platform without oversight can be "weaponized" by various actors, but a completely closed one is weaponized by the state. There is no easy solution, but one potential path for Kazakhstan is the development of indigenous platform governance capacity: e.g., Kazakh-language content moderation algorithms and fact-checking systems that can quickly counter fake content in Kazakh. Presently, local civil society and initiatives like StopFake.kz are the main counter to falsehoods. If Kazakhstan wants digital sovereignty *without* resorting to broad censorship, investing in algorithmic tools (perhaps AI that detects coordinated activity or deepfake videos) is essential. Encouragingly, the step to monitor and publicize disinformation campaigns (as TSARKA did in 2024) indicates a move toward transparency rather than blanket blocking, which is a healthier approach for discourse.

However, the risk is that calls for digital sovereignty can justify overreach. For instance, demanding access to Telegram's data or putting pressure on tech companies (Haidar, 2025) might curb freedom of expression if abused. Media control in Kazakhstan might shift from overt to algorithmic – instead of jailing journalists (which still happens, sadly), the state might rely on keeping unfavorable content algorithmically out of sight. The conclusion from our research is that such algorithmic suppression often eventually comes to light (users notice, leaks happen, as with Twitter's internal biases revelation). Therefore, a *non-obvious but critical point* is that algorithmic transparency and accountability should be part of any nation's quest for digital sovereignty. If algorithms are discourse subjects, perhaps they need to be treated akin to media entities – with responsibilities and oversight. Some scholars argue for algorithmic audits and even the idea that users have a right to know why they see what they see (Pasquale, 2015). Implementing this in authoritarian contexts is tricky, as those regimes benefit from algorithmic opaqueness. Yet, Kazakhstan's relatively more pluralistic society (compared to Russia) could push for at least *some* transparency (for example, requiring platforms to report the most trending content and its origin during election periods, which could expose bot-driven trends).

3. The strategies bots use—code-switching, mimicking dialects, etc.—have sociolinguistic consequences. One could argue they contribute to language change or at least to perceptions of language use. If people keep encountering badly written Kazakh from bots, does it affect how they perceive written Kazakh online? Possibly, it could either degrade trust in online commenters generally or stigmatize certain styles as "bot-like." There are anecdotes from our focus group interviews (we held a few informal chats with young Kazakh internet users) where participants said things like: *"When I see someone writing only in formal Kazakh on Facebook praising the akim (governor), I immediately think, ignore it, it's a bot or paid."* Essentially, a new digital literacy has formed: *the ability to read linguistic clues to judge authenticity*. This is promising—users are adapting by developing what we might call algorithms of the mind to filter information ("if too formulaic, likely fake"). But not everyone has that literacy; less savvy or older users might take things at face value. That's why many propaganda and scam messages target older people (e.g., in Russian WhatsApp or Telegram circles).

The multilingual tactics also speak to language politics. The Kazakh government's promotion of Kazakh could inadvertently make Kazakh the preferred vehicle for propaganda as well, since it's seen as patriotic. We noted how some propaganda channels switched to more Kazakh posts post-2022. This is a double-edged sword: on one hand, more content in Kazakh (even if propaganda) normalizes using the language online – potentially a win for language revitalization; on the other hand, if that content is manipulative, it could taint online Kazakh discourse with suspicion. Language ideology theory (Irvine & Gal, 2000) suggests that languages often get ideologically tied to morality or truth. There is a scenario where people might say "I only trust news if it's in Kazakh" or conversely "pro-gov propaganda is always in Kazakh these days, I trust Russian sources more." Either outcome is problematic if driven by bot proliferation rather than genuine community voices. It underscores that *linguistic diversity online needs support from authentic content creation*, to drown

out inauthentic noise. Encouraging more real users to engage in Kazakh (and in reasoned debate) can make it harder for bots to dominate narrative without standing out.

4. Our findings reflect on how platform companies are responding. In the English sphere, Twitter and Facebook have taken down large bot networks (e.g., Twitter's bans of tens of thousands of IRA-linked accounts, Facebook's removal of the Kazakh NSC-linked network in 2020) (Freedom House, 2022). These actions demonstrate the platforms' algorithms and human teams trying to cleanse discourse. Yet, in non-English contexts, responses have been slower or less effective, partly due to lack of linguistic expertise. The fact that we still saw plenty of obvious bots on Kazakh pages in 2022–23 suggests that Facebook, for instance, doesn't prioritize Kazakh-language moderation. This is a common criticism (Global Voices, 2022): moderation outside English is often weaker, giving manipulators a free hand. As a result, the heavy lifting of debunking or calling out bots fell to local initiatives or researchers.

Given this, one recommendation is that platforms invest in local partnerships – e.g., with Kazakh universities or tech companies – to improve their detection algorithms for local languages and contexts. If an algorithm can be trained to recognize Nurbot patterns (e.g., an account that only comments on political posts with known pro-government phrases), it could automatically downrank or flag them. Similarly, detection of code-switched malicious content is a technical frontier: it requires understanding two languages at once. Advances in multilingual AI (like transformer models that cover 100+ languages) may help here. Already, projects like Multilingual BotBuster (Ng et al., 2023) aim to handle this, and their results (bots in every country ~20% of content) show it's feasible to map the bot landscape broadly.

5. The cumulative impact of these manipulations is a coarsening and fragmentation of public discourse. We see that *in all three cases, trust in online information is eroded*. Americans increasingly suspect each other of being bots or influenced by algorithms (e.g., the phrase "do you live in a filter bubble?" entered common parlance). Russians have internalized that many glowing comments about the government are fake, but that also leads to cynicism – they might then also dismiss genuine grassroots criticism as possibly orchestrated (the logic of "info chaos" as mentioned by a Diplomat piece (Keegan, 2022) on Kazakhstan's info space during crisis: when you don't know what's real, it paralyses collective action). In Kazakhstan, the government urging people to rely only on "official sources" (BBC Monitoring, 2024) could further centralize information flows, ironically playing into the hands of a controlled narrative. If people conclude "social media is full of lies and bots, I'll just believe the government TV," that's a victory for authoritarian control but a loss for a pluralistic discourse.

However, there's also a counter-trend: digital literacy growth. Especially among youth, there's a keen awareness and even humor about bots ("lol that comment is such a bot thing to say"). This awareness is the first step in resilience. Our results underscore many "non-obvious" strategies precisely to inform educators and the public about them. Once a strategy is recognized broadly (say, people learn that mixed-language posts might be suspect), manipulators lose that tool or have to invent new ones. Essentially, there is an evolutionary pressure in the discourse environment – with each exposure and debunking, the ecosystem adapts. The downside is, new strategies may be even harder to detect (perhaps deepfake videos with fluent Kazakh speech by AI voices could be the next wave, to impersonate figures or eyewitnesses).

6. Treating algorithms as discourse subjects means acknowledging their agency-like influence. Should they then be accountable like subjects? This is a theoretical but also practical question. For example, Twitter's algorithm favoring certain political content – should that be considered an editorial act (and thus Twitter be seen as a publisher in that regard)? If so, it might conflict with how these companies claim neutrality. In Russia, they've basically forced algorithms to be not neutral – VK and Yandex news had to demote content from undesirable sources by law or guidance. That is an overt harnessing of algorithm as subject under state control. In open societies, the question is whether algorithms can be governed in a way that *enhances* discourse quality rather than undermines it. Some proposals include algorithmic audits by independent agencies, user choice in algorithm (e.g., chronological vs. ranked feed toggles), or even public service algorithms for important information.

For Kazakhstan and similar states, one could imagine a positive use of algorithms: say, during emergencies, an algorithm that pushes verified information (like safety instructions) in Kazakh and Russian to all users on multiple platforms – essentially co-opting the virality for good. There were small examples: during COVID-19, some platforms pinned WHO or local health info at the top of feeds. Extending that concept, algorithmic governance could be used to *counter*manipulation by flooding with facts. But that requires trust and competence; if the government simply uses it to flood propaganda, it's no better than bots. It would need a coalition of credible voices, perhaps leveraging algorithms to amplify corrective discourse.

Finally, the interplay of media control and user rights emerges strongly. Freedom of expression is at stake when addressing these manipulations. Tighter control might reduce fake content but also silence dissent. The art is in distinguishing the two, which often requires human judgment and rule of law. Purely algorithmic solutions might accidentally censor activists thinking they are bots (false positives), or conversely, whitelisted government bots might skate through. Thus, multi-stakeholder oversight (involving tech experts, civil society, and some form of judicial review) could be necessary when implementing measures to curb algorithmic manipulation.

One non-obvious conclusion from this research is that *linguistic diversity can be both a vulnerability and a strength* in combating algorithmic manipulation. It's a vulnerability because as we saw, bots exploit multiple languages to increase reach and avoid detection. But it can be a strength because a diverse linguistic environment means manipulators must split efforts, and truly mastering multiple languages to a native level (to fool people) is hard for them. Moreover, citizens

who operate in two languages might access information from two sources (e.g., a Kazakh-speaker might also know what Russian media is saying, providing a form of comparison that monolingual audiences don't get). In Kazakhstan, many people could see both the Russian narrative and the Kazakh official narrative and recognize discrepancies. That bilingual perspective could foster a more skeptical, analytical mindset, ironically immunizing them better than if they only heard one language narrative. This is a nuanced point that complicates the assumption that linguistic fragmentation always harms the public sphere; in some cases, it provides an alternate viewpoint. The implication is that efforts to counter manipulation might leverage bilingual education – encouraging citizens to cross-check info in both languages they understand.

Another subtle finding is the emergent norm of calling out bots (audience resistance). In all three contexts, segments of the public are actively participating in cleaning the discourse: Americans creating bots to counter-troll Russian trolls (there was an initiative where people launched positive bots to flood out hate hashtags), Russians ironically using memes to mock obvious propaganda ("Lake Troll" memes, etc.), and Kazakhs labeling comments as "Nurbot" to discredit them. Such organic corrective behavior is a form of community self-regulation.

V. Conclusion

Information operations have evolved beyond crude spam into subtle, linguistically strategic interventions. Disinformation actors now mimic dialects, code-switch fluidly, and craft culturally embedded personas that evade simple detection. In Kazakhstan, this includes hybrid-language bots and semi-authentic comment strategies that reflect a high degree of localization. Such tactics challenge both human and automated systems, requiring more nuanced, context-aware responses.

States have responded with varying models of digital control. Russia has moved toward algorithmic centralization and narrative consolidation, while Kazakhstan maintains a looser, more pluralistic space—albeit with vulnerabilities to external manipulation. The Kazakh experience in 2024, involving both public disinformation countermeasures and civic engagement, suggests a potential third model: responsive, open, and locally anchored.

A key insight is the central role of users. Digitally literate individuals, able to identify manipulation and navigate discursive traps, function as decentralized moderators. Their interpretive agency complements institutional responses and offers a pathway toward greater resilience. Platforms could enhance this role by designing affordances that promote context, contradiction, and critical engagement.

Theoretically, the findings bridge discourse analysis and algorithm studies, arguing for recognition of algorithms as actors embedded in linguistic economies. This reframes traditional sociolinguistic questions—such as who is heard, what is seen as "proper" language, and how norms shift under digital pressure. Future research might track algorithmic influence longitudinally or examine parallel dynamics in other multilingual societies.

The digital public sphere of 2025 is not a neutral space but a contested terrain where language, power, and computation converge. Kazakhstan's case shows that even states with modest digital infrastructure are caught in this struggle—and can, through innovation and vigilance, model responses others may follow. Ultimately, ensuring that algorithms support democratic dialogue rather than distort it requires systemic rethinking across legal, technological, and cultural domains.

References

[1] Aksholakova, A., & Ismailova, R. (2013). The language policy of Kazakhstan and the state program of functioning and development of languages for 2011–2020. *Procedia – Social and Behavioral Sciences*, *93*, 1580–1586. https://doi.org/10.1016/j.sbspro.2013.10.121

[2] BBC Monitoring. (2024). *Kazakhstan and regional social media monitoring reports*.

[3] Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.

[4] Bradshaw, S., Bailey, H., & Howard, P. N. (2021). *Industrialized disinformation: 2020 global inventory of organized social media manipulation*. Computational Propaganda Research Project. https://comprop.oii.ox.ac.uk

[5] Cadwalladr, C. (2017, March 17). *The Cambridge Analytica files*. The Guardian.

[6] Danaher, J. (2016). The threat of algocracy: Reality, resistance and accommodation. *Philosophy & Technology*, *29*(3), 245–268. https://doi.org/10.1007/s13347-016-0221-6

[7] DataReportal. (2023). *Digital 2023: Kazakhstan*. We Are Social & Meltwater.

[8] Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, *59*(7), 96–104. https://doi.org/10.1145/2818717

[9] Freedom House. (2022). *Freedom on the Net 2022: Kazakhstan*.

[10] Global Voices. (2022). *Social media moderation in non-English languages: Gaps and challenges*. https://globalvoices.org

[11] Howard, P. N., Ganesh, B., Liotsiou, D., Kelly, J., & François, C. (2018). *The IRA, social media, and political polarization in the United States, 2012–2018*. Computational Propaganda Research Project. https://comprop.oii.ox.ac.uk

[12] Huszár, F., Kveton, B., Taly, A. et al. (2022). Algorithmic amplification of political content on social media. In *Proceedings of the 38th Conference on Neural Information Processing Systems*.

[13] Irvine, J. T., & Gal, S. (2000). *Language ideology and linguistic differentiation*. In P. Kroskrity (Ed.), *Regimes of language: Ideologies, polities, and identities* (pp. 35–83). School of American Research Press.

[14] Kamalova, A. (2025). Bilingualism and online political discourse in Kazakhstan. *Central Asian Communication Studies*, *7*(1), 12–29.

[15] Kozhanova, A. (2019). Digital propaganda and the russophone online space. *Journal of Eurasian Studies*, *10*(2), 45–61.
[16] McCombs, M., & Shaw, D. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, *36*(2), 176–187.
[17] Ng, H., et al. (2023). Multilingual BotBuster: Detecting coordinated inauthentic behavior across languages. *AI & Society*, *38*, 567–584.
[18] Nocetti, J. (2015). Algorithmic governance and the shaping of digital discourse. *Journal of Digital Policy*, *2*(1), 44–61.
[19] Ospanova, K. (2020). Astroturfing and digital manipulation in Kazakhstan. *Central Asian Journal of Media Studies*, *5*(2), 77–93.
[20] Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
[21] Reddit. (2021). *User discussions referencing Kazakhstan census language-use data*.
[22] Sapakova, A., Tursunov, M., & Imanbekov, B. (2024). *Social media trends and digital engagement in Kazakhstan*. Almaty: Central Asia Media Analytics.
[23] Sánchez, M. (2023). Disinformation on Telegram: Case studies from Eurasia. *Journal of Online Misinformation*, *3*(1), 12–35.
[24] Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, *23*(3–4), 193–229. https://doi.org/10.1016/S0271-5309(03)00013-2
[25] TSARKA via BaigeNews. (2024, December 15). *Anomalous bot activity around TikTok protest video*. BaigeNews.
[26] Twitter Safety. (2020). *Quarterly enforcement report: April–June 2020*. https://transparency.twitter.com/en/reports

Saltanat Beissembayeva, a candidate of Philology Sciences, is currently head of the Department of Russian Language and Literature at Khoja Ahmed Yasawi International Kazakh-Turkish University. She has published in peer-reviewed international journals. Author of educational and methodical manuals. Her research focuses on paralinguistics, intercultural communication, language theory and teaching methods. ORCID iD: 0000-0002-4431-7558. E-mail address: saltanat_rb@mail.ru

Marzhan Zhylkybayeva, a candidate of Philology Sciences, is currently an associate professor at the Department of Kazakh Philology at Caspian University of Technology and Engineering named after Sh.Yesenova. She has published her work in international conferences and in journals recommended by the CCFES of the Ministry of Education and Science of the Republic of Kazakhstan. She is also the co-author of the educational and methodological manual "Language manual for preparing for the exam to assess the level of proficiency in the Kazakh language by testing", developed based on the order of the Committee on Languages of the Ministry of Culture of the Republic of Kazakhstan. Her research focuses on areal linguistics, ethno-linguistics and the problems of the modern Kazakh language. Currently, she works as a researcher in the research group of the program-targeted financing project (2024-2026) "Ancient Turkic place names in the sources of China and other Eastern countries - one of the foundations of modern Turkic integration - historical, etymological, linguistic approach". ORCID iD: https://orcid.org/0000-0002-2329-6065. E-mail address: sarsekeshovna@mail.ru

Gulzhan Seydullaeva (corresponding author), a candidate of Philology Sciences, is currently an associate professor at the Department of Social Sciences and Humanities of the International Educational Corporation (KazHACA) in Kazakhstan. The author of articles in peer-reviewed international journals, as well as the author of the monograph "Communicative Linguistics", textbooks on the Russian language for students. The winner of the Bolashak international internship within the framework of the 500 Scientists program, Turkey, Ankara University. Khoja Bayram was conducted. Her research focuses on corpus linguistics, terminology and stylistics. ORCID iD: https://orcid.org/0000-0002-2437-9321. E-mail address: g.a.seidullaeva@mail.ru

Karlygash Aubakirova, a doctor of Philosophy (PhD), literary critic-scientist. Teacher of the Department of "Kazakh philology and journalism", Shakarim University, senior researcher of the research center "Abay and national rukhaniyat". Conducts research work in the field of fiction and censorship, Kazakh literature of the early twentieth century, Alash studies, translation. He is the author of the monograph "Alash translation (based on the translations of Zhusupbek Aimauytuly)" and many research articles. ORCID iD: 0000-0002-1872-9591. E-mail: kz.aubakirova@gmail.com

Moldir Amangazykyzy, PhD, is currently an Associate Professor at A.K. Kussayinov Eurasian Humanities Institute, Astana, Kazakhstan. Her research in Kazakh literary studies focuses on literary representations of urban themes, where she has pioneered the identification and analysis of city models in contemporary Kazakh prose. In her academic works, she has systematically traced the transformation of urban imagery through various historical periods, from ancient Turkic monuments and medieval literature to 20th-century texts and contemporary Kazakh narratives. ORCID iD: 0000-0001-8027-0673. E-mail: moldir_amangazykyzy@mail.ru

Ryskul Akniyazova is a teacher of Kazakh language and literature, a master teacher. She was awarded the medal "Y.Altynsarin". Works as a teacher of Kazakh language and literature at KSU "Akshukur Lyceum school", Tupkaragan district, Mangystau region. Participated in international and national scientific and practical conferences and published articles. Author of the methodical manual "Abish studies" for teachers. Winner of the republican stage of the contests "The best author's program" and "Festival of Pedagogical Ideas". Her students have repeatedly won prizes at the republican stage of scientific projects and creative competitions. ORCID iD: https://orcid.org/0009-0008-2881-9380. E-mail address: aiakoz00@mail.ru