Multiscript Language Technologies and Digital Inclusion: A Sociotechnical Study of Kazakh Script Conversion

Aigerim Sadykova[1,*] Nurlan Beketov[2], Madina Tursynbayeva[3]

[1] Department of Information Systems, M. Auezov South Kazakhstan University, Shymkent, Kazakhstan
[2] Department of Computer Science, Korkyt Ata Kyzylorda University, Kyzylorda, Kazakhstan
[3] Department of Sociology and Social Work, Shakarim University of Semey, Semey, Kazakhstan

* Email: *a.sadykova@auezov.edu.kz (Corresponding Author)*

Abstract

Kazakh is a major Turkic language written and read through Arabic-based, Cyrillic-based, and Latin-based scripts across different communities, regions, archives, and digital platforms. This multiscript condition is often treated as a narrow technical problem of transliteration accuracy, yet it also shapes who can search for public information, access education, preserve family records, participate in e-government, and maintain linguistic identity in data-driven societies. This paper develops a sociotechnical study of Kazakh script conversion by connecting neural conversion methods, loanword-aware prompting, corpus governance, and digital inclusion. Building on a secondary analysis of recent Kazakh multiscript conversion benchmarks, the study reinterprets character error rate (CER) and word error rate (WER) as proxies for accessibility friction, institutional reliability, and cultural continuity. The analysis shows that prompt-constrained Transformer conversion substantially reduces word-level friction across six conversion directions but also reveals that model accuracy alone is insufficient for inclusive deployment. Script conversion systems affect people through interface design, standardization choices, metadata practices, education policies, data rights, and community trust. The paper contributes a multilayer framework that links script ecology, linguistic resources, conversion services, access settings, governance arrangements, and inclusion outcomes. It further proposes design principles for transparent, auditable, and community-sensitive Kazakh language technologies. The findings suggest that multiscript conversion should be governed not merely as an automation service but as digital public infrastructure for linguistic equity.

Keywords: Kazakh; multiscript conversion; digital inclusion; sociotechnical systems; language technology; loanword prompts; Transformer

Article History
Received: April 24, 2025
Revised: June 12, 2025
Accepted: August 19, 2025
Available Online: September 30, 2025

Multiscript Language Technologies and Digital Inclusion: A Sociotechnical Study of Kazakh Script Conversion

1. Introduction

Language technologies increasingly mediate access to education, public services, cultural memory, and economic participation. Search engines, digital libraries, administrative portals, social media platforms, and learning applications do not merely store written language; they classify it, index it, normalize it, convert it, and make it discoverable or invisible. For languages written in one stable script, these processes are already politically and technically consequential. For languages distributed across multiple scripts, the consequences become more complex because the same word, person, place, or institution may exist in several script forms that are not equally supported by software systems, keyboards, fonts, search algorithms, or government databases. This infrastructure perspective treats language tools as part of access systems rather than as isolated software components (Star and Ruhleder,1996). The same point is visible in attention-based modeling, where technical architecture becomes meaningful only when matched to data and task context (Bahdanau et al.,2015).

Kazakh is a particularly important case for studying this problem. It is used across Kazakhstan, China, Russia, Mongolia, and transnational communities, and it has long circulated in Arabic-based, Cyrillic-based, and Latin-based writing systems. These scripts are not merely interchangeable visual encodings. They are associated with migration histories, schooling systems, political reforms, publishing infrastructures, religious and cultural memories, and cross-border communication patterns. A Kazakh speaker who reads one script fluently may not be equally comfortable with another. A family record, school textbook, regional newspaper, or public notice may become less accessible when a platform supports only one script variant. In this sense, script conversion is not only a natural language processing task; it is also a question of digital inclusion. The risk is sharper for languages whose online presence is uneven across scripts and regions (Kornai,2013).

Technical literature has made significant progress in modeling Kazakh script conversion. Earlier systems relied on rule tables, dictionaries, or statistical learning. More recent systems use neural sequence-to-sequence models, attention mechanisms, and prompt-guided learning. The source manuscript motivating this study proposes a multi-script neural machine conversion method incorporating loanword prompts and reports strong gains overrule-based, dictionary-based, and RNN-based baselines. Its central technical insight is that loanwords, regional vocabulary, vowel harmony, consonant alternation, and morphologically rich word forms create ambiguity that cannot be solved by simple letter mapping alone. A Transformer architecture guided by loanword prompt data improves conversion quality because it combines contextual modeling with explicit linguistic constraints. Prior work on Arabic-Kazakh character processing shows that script conversion depends on encoding, font behavior, and orthographic convention, not only linguistic mapping (Dong et al.,2018). Adapter-based multilingual methods also show why small task-specific adjustments can matter for low-resource deployment (Bapna and Firat,2019).

This paper takes technical development as a starting point and asks a broader question: what does script conversion mean for digital inclusion? In many technological assessments, a lower character error rate is treated as an endpoint. In actual use, however, a small error may misidentify a person, hide a document from search results, corrupt a technical term, or weaken trust in a government or education platform. Conversely, a technically strong model may still be exclusionary if it is deployed without transparent standards, community review, offline access, support for older publications, or safeguards against dialectal erasure. The issue is therefore sociotechnical: digital inclusion emerges from the interaction between algorithms, data, interfaces, institutions, and users. This wider framing is consistent with transliteration research that treats script transfer as a language-technology problem with user-facing consequences (Karimi et al.,2011). It also connects to broader discussions of artificial intelligence as a socio-organizational technology rather than a purely computational artifact (Lu,2019).

The paper makes three contributions. First, it reframes Kazakh multiscript conversion as a sociotechnical infrastructure problem rather than a narrow transliteration task. Second, it conducts a secondary analysis of benchmark results across Arabic, Cyrillic, and Latin conversion directions and interprets error reductions as reductions in access friction. Third, it proposes practical design and governance principles for inclusive multiscript language technologies, including loanword

See: https://inatgi.in/index.php/jtis/index for more information. https://doi.org/10.63646/jtis.2025.030303

transparency, corpus stewardship, metadata interoperability, user choice, community validation, and fairness-oriented evaluation. The purpose is not to replace technical metrics, but to connect them with the social conditions under which language technologies become trustworthy and useful. Named-entity transliteration research further supports the need to protect personal names and domain-specific terms during conversion (Al-Onaizan and Knight,2002).

The remainder of the paper is organized as follows. Section 2 reviews relevant research on multiscript language processing, low-resource NLP, and sociotechnical digital inclusion. Section 3 describes the Kazakh multiscript context and the specific conversion problems that matter for access. Section 4 presents the research design and analytical framework. Section 5 provides the technical and inclusion-oriented analysis of conversion performance. Section 6 discusses deployment implications for education, public services, cultural heritage, and platform governance. Section 7 offers design recommendations for inclusive Kazakh language technologies. Section 8 concludes the paper and identifies future research directions. Kazakh corpus-development studies show that reusable language resources are central to sustainable language technology ecosystems (Akhmed-Zaki et al.,2021).

2. Literature Review and Theoretical Background

2.1 Multiscript Language Processing

Multiscript language processing refers to computational methods that handle languages written in more than one script or orthographic system. The problem appears in South Asian languages, Mongolic languages, Turkic languages, Chinese romanization, Arabic-script adaptations, and many indigenous or minority-language contexts. In some cases, script conversion is close to transliteration, where a source character or phoneme can be mapped to a target symbol. In other cases, conversion requires morphological analysis, contextual disambiguation, lexical substitution, or standardization of borrowed terms. Kazakh belongs to the second category because its three major scripts differ in alphabet size, phonological representation, and regional usage. A modular transliteration framework is useful in this setting because it allows rules, learned models, and language-specific resources to be combined (Kumaran and Kellner,2007).

Classic script conversion systems were attractive because they were explainable and easy to implement. A table of character correspondences could be encoded in software, and predictable sound changes could be handled by handcrafted rules. Such systems remain valuable in low-resource settings because they require fewer training data and can be maintained by linguists. However, they struggle with ambiguous correspondences, irregular loanwords, and contextual forms. Dictionary-based systems improve known-word accuracy but fail when words are absent from the lexicon, when borrowed vocabulary changes across regions, or when a stem appears with many affixes. Statistical methods can learn patterns from corpora, but their performance depends heavily on the quantity and balance of aligned data. Early statistical transliteration work demonstrated that cross-script mapping must account for both pronunciation and spelling variation (Knight and Graehl,1998).

Neural methods have changed the conversion landscape by treating script conversion as a sequence modeling problem. Recurrent networks and attention-based models can learn context-sensitive mappings that are not easily expressed as rules. Transformer architecture further improves the modeling of long-range dependencies through self-attention [Vaswani et al., 2017]. Yet neural methods also introduce new risks. They may produce fluent but incorrect forms, overfit dominant regional spellings, or obscure the reason for a conversion decision. Prompt-guided conversion and lexicon-aware attention provide a middle path by combining neural flexibility with explicit linguistic knowledge. This is especially useful for loanwords, names, technical terms, and other forms in which social usage and phonological rules do not fully align. Subword modeling is especially relevant because rare words, inflected forms, and loanwords often determine conversion quality in low-resource settings (Sennrich et al.,2016). Encoder-decoder models also provide a basis for learning context-sensitive sequence transformations instead of relying on fixed lookup tables (Cho et al.,2014). Recent information-system studies similarly highlight the need to align artificial intelligence models with practical deployment conditions (Zhang and Lu,2021).

2.2 Digital Inclusion and Language Technology

Digital inclusion research emphasizes that access is not limited to physical connectivity. People need devices, affordable networks, digital literacy, meaningful content, accessible interfaces, institutional support, and the ability to use technologies in languages that matter to them [Warschauer, 2004]. When a community's script is unsupported, poorly rendered, or inconsistently indexed, users may technically have internet access while remaining excluded from search, education, and public communication. Script support therefore functions as an infrastructural condition for inclusion. Digital inclusion therefore depends on skills, meaningful use, and local content in addition to network access (Hargittai,2002). This point is consistent with critiques that the digital divide should not be reduced to a binary distinction between users and nonusers (Selwyn,2004).

The politics of language technology are often hidden because software presents normalization as a neutral technical operation. However, every conversion standard privilege certain spelling, phonological interpretations, or institutional choices. A search engine may decide that two script forms are equivalent, that one form is canonical, or that a user must enter text in a dominant script to retrieve results. A school platform may convert learning materials automatically, but it may not explain how technical terms or personal names are handled. An archive may digitize older newspapers without building cross-script metadata. These choices influence whose knowledge becomes searchable and whose language practices are treated as exceptions. A corresponding fields view of digital exclusion helps explain why linguistic access, institutional participation, and service outcomes reinforce one another (Helsper,2012). Usage-based digital divides also show why a converter must be evaluated through actual tasks rather than download counts alone (van Deursen and van Dijk,2014).

Sociotechnical theory provides useful concepts for understanding these dynamics. Infrastructure is not only a physical or computational system; it is embedded in social practices, standards, work routines, and communities of maintenance [Star and Ruhleder, 1996]. Technological artifacts also embody institutional values and political arrangements [Winner, 1980]. Applied to Kazakh conversion, this means that a converter is not simply an input-output tool. It is part of a larger chain that includes keyboard layouts, OCR systems, fonts, corpora, editorial standards, user training, archival metadata, government forms, and quality assurance practices. A converter with high average accuracy may still fail as inclusive infrastructure if it is inaccessible, opaque, or socially misaligned. Systematic reviews of digital inequality further suggest that skills, usage, and outcomes must be assessed together (Scheerder et al.,2017). A practice-based view of technology clarifies why conversion tools reshape routines while also being reshaped by local users (Orlikowski,1992). Digital transformation research reinforces this need to study technical change together with organizational and social adaptation (Lu,2025).

2.3 Low-Resource NLP and Responsible Evaluation

Kazakh is often described as a low-resource language in NLP because high-quality annotated corpora, parallel datasets, speech resources, and standardized benchmarks remain limited when compared with English, Chinese, Russian, or other high-resource languages. Low-resource conditions amplify the importance of data governance. A small corpus may overrepresent urban media language and underrepresent rural varieties, diaspora usage, or older scripts. Loanword dictionaries may encode editorial choices that are useful for conversion but contested by users. Evaluation sets may reward a single target spelling while penalizing legitimate alternatives. Global NLP benchmarking shows that linguistic diversity remains unevenly represented in mainstream datasets and models (Joshi et al.,2020). Low-resource NMT surveys also show that data scarcity, domain mismatch, and evaluation instability are persistent barriers (Ranathunga et al.,2023).

Responsible evaluation therefore requires more than aggregate accuracy. Character error rate and word error rate provide essential diagnostics, but they do not capture whose words are mis converted, which domains suffer most, whether names are protected, or whether users can correct system output. Bias evaluation in NLP has shown that aggregate metrics can hide systematic errors affecting marginalized groups [Blodgett et al., 2020]. For multiscript language technologies, the parallel concern is that a model may perform well on news text while failing on family names, regional expressions, educational terms, or cultural heritage documents. This paper accordingly treats technical accuracy as one layer of analysis and connects it to user-facing inclusion outcomes. Low-resource translation research recommends combining automatic metrics with careful qualitative analysis and task-level interpretation (Haddow et al.,2022). Significance testing is also important because small numerical differences can be misleading in language-technology evaluation (Dror et al.,2018).

3. Kazakh Multiscript Context and Inclusion Problem

Kazakh script diversity has both linguistic and social dimensions. The Arabic-based script is used by many Kazakh communities in Xinjiang and appears in regional media, cultural publications, personal writing, and historical materials. The Cyrillic-based script has been widely used in Kazakhstan and post-Soviet institutions. Latin-based forms are associated with modernization, digital internationalization, and ongoing language reform. These scripts share a relationship to the same language, but their practical use is mediated by different schooling histories, publishing infrastructures, software defaults, and cross-border communication needs. Research on Indigenous and minoritized language technologies similarly shows that digital tools can support revitalization only when communities retain agency (Galla,2016).

The conversion problem is challenging because script systems differ in how they represent vowels, consonants, and borrowed words. Kazakh vowel harmony interacts with script-specific orthographic conventions. In Arabic-based Kazakh, soft-sign behavior and the placement of front-vowel markers can change how a word is written. Consonant mutation occurs when suffixes are added to stems, making it necessary to understand morphology rather than simply map final letters. Loanwords further complicate conversion because they often enter Kazakh through different contact languages, including Russian, Arabic, Chinese, and English. Their spellings can preserve pronunciation, borrow regional conventions, or diverge from vowel harmony rules. Kazakh handwritten text research also illustrates that script processing often requires attention to visual form, segmentation, and noisy real-world documents (Toiganbayeva et al.,2022).

These linguistic features become inclusion issues when digital systems treat conversion as a mechanical substitution task. A student who receives a converted science text may encounter corrupted terminology. A parent trying to search for school information may not retrieve relevant pages because the same term is indexed in another script. A public office may store a name in Cyrillic while a citizen's family documents use Arabic-based spelling. A cultural archive may digitize materials but fail to connect them to modern query forms. In each case, conversion errors create practical barriers to participation. Kazakh parallel-corpus development further indicates that domain-balanced data are essential for reliable multilingual and multiscript applications (Yeshpanov et al.,2024).

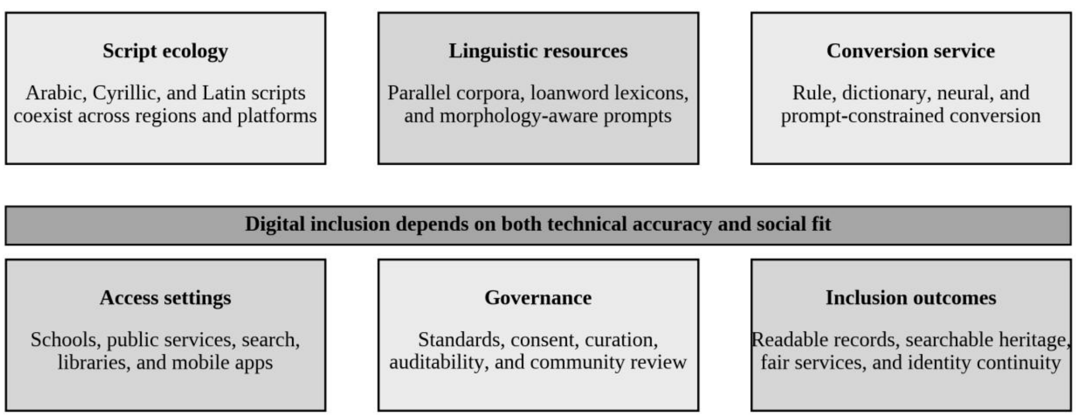

Figure 1. Sociotechnical dimensions of Kazakh multiscript conversion for digital inclusion

Figure 1 summarizes the core argument. Script ecology, linguistic resources, conversion services, access settings, governance, and inclusion outcomes are mutually dependent. The technical service occupies only one part of the system. The quality of conversion depends on corpora and lexicons, while its social value depends on where it is used and how people contest or correct the output. This structure also explains why sociotechnical study is needed. Without a social layer, script conversion is evaluated as an isolated model. Without a technical layer, inclusion policy remains disconnected from the actual error patterns that shape user experience. Recent reviews of Central Asian Turkic NLP emphasize that Kazakh, Uzbek, Kyrgyz, and Turkmen share low-resource challenges while requiring language-specific solutions (Veitsman and Hartmann,2025).

Table 1. Kazakh script conversion as a sociotechnical problem

Technical issue	Linguistic source	User-facing risk	Inclusive design response
Many-to-one letter mappings	Different alphabet inventories across scripts	Search misses, name inconsistency, and wrong entity linking	Cross-script indexing with uncertainty display
Vowel harmony and reduction	Front/back vowel behavior and soft-sign conventions	Unreadable or unnatural converted text in learning materials	Morphology-aware conversion and domain testing
Consonant mutation	Suffix-driven stem alternation in agglutinative morphology	Incorrect stems in administrative or educational text	Stem-aware prompt construction and correction logging
Loanword variation	Borrowings from Russian, Arabic, Chinese, English, and other sources	Ambiguity in technical terms, names, and public information	Curated loanword lexicons with transparent alternatives
Regional spelling practices	Cross-border media and community usage differences	Perceived erasure of local language identity	User-selectable regional profiles and community review

Table 1 shows that a technical error is not socially uniform. Errors in a casual social media sentence may be inconvenient, but errors in a legal name, medical instruction, cultural archive label, or school examination item can have more serious consequences. An inclusive converter should therefore be evaluated by domain, user group, and communicative context. This does not mean that every variant must be accepted without standardization. Rather, the system should make standardization explicit, allow user control where appropriate, and preserve enough metadata to support traceability. Benchmarking work on Turkic language transfer also shows that performance varies across related languages and tasks, which limits one-size-fits-all evaluation (Senel et al.,2024).

4. Research Design and Analytical Framework

This study uses a mixed conceptual and secondary quantitative design. The conceptual component develops a sociotechnical framework for analyzing Kazakh script conversion as digital inclusion infrastructure. The secondary quantitative component reanalyzes benchmark results from a recent Kazakh multiscript conversion study that compared rule-based, dictionary fusion, RNN, RNN-Attention, and loanword-prompt neural conversion methods across six conversion directions. The original technical study measured character error rate (CER) and word error rate (WER). The present paper does not duplicate its model design. Instead, it interprets the same performance signals as indicators of conversion friction in social settings. A practice lens is useful here because it connects technical artifacts with routines, institutions, and situated use (Orlikowski,2000).

The analytical framework has three layers. The first layer is technical reliability, measured through error rates, directionality, and improvement over baselines. This layer asks whether the system produces the intended target script with sufficient accuracy. The second layer is inclusion relevance, which interprets error as a barrier to search, readability, institutional reliability, or identity continuity. This layer asks who is affected by errors and in which settings. The third layer is governance readiness, which considers whether the system supports transparency, correction, auditability, data stewardship, and community participation. This layer asks whether users and institutions can trust the conversion infrastructure over time. Sociotechnical transition theory similarly explains why technical improvements become durable only when they fit institutional rules and user practices (Geels,2004).

The use of secondary benchmark data is appropriate for this study because the central research question concerns interpretation and deployment rather than model training. Technical benchmarks can reveal where conversion is more or less reliable. Sociotechnical analysis can then explain why these differences matter. For example, a lower word error rate in Cyrillic-to-Latin conversion may indicate that modern digital services can support Latin reform materials more effectively, while higher error rates in Arabic-to-Cyrillic conversion may signal greater risk for cross-border archives and Xinjiang-Kazakhstan information exchange. The same numerical result can thus be read as both a model property and an inclusion signal. Transition studies also warn that evaluation should consider regimes, infrastructures, and stakeholder incentives rather than isolated devices (Markard et al.,2012).

The study also uses a scenario-based lens. Four access settings are considered: education, public services, cultural heritage, and platform search. These settings were selected because they represent common points at which script conversion shapes participation. Education involves textbooks, assignments, terminology, and student identity. Public services involve names, forms, notices, and official records. Cultural heritage involves older publications, oral history transcripts, family documents, and museum metadata. Platform search involves user queries, recommendation systems, and cross-script retrieval. Across these settings, the paper examines how conversion accuracy, explainability, and governance combine to influence inclusion. Model-reporting practices offer a useful template for documenting limitations, intended uses, and performance differences across user groups (Mitchell et al.,2019). Research on secure information systems further supports documenting operational risks when digital tools enter public-service workflows (Lu and Xu,2019).

Table 2. Evaluation layers for inclusive multiscript conversion

Layer	Core question	Indicative measures	Why it matters
Technical reliability	Does the converter produce accurate target-script text?	CER, WER, direction-level error, loanword accuracy	Prevents wrong names, terms, and sentences
Access relevance	Does the output support real user tasks?	Readability, search recall, domain-specific error, correction burden	Connects accuracy to education, services, and archives
Governance readiness	Can decisions be inspected, corrected, and maintained?	Versioning, logs, corpus documentation, community review	Builds trust and long-term accountability
Equity sensitivity	Whose language practices are supported or ignored?	Regional variants, dialectal coverage, name protection, offline access	Avoids turning standardization into exclusion

Table 2 operationalizes the study's evaluation logic. The layers are not alternatives; they are cumulative. A system that fails in technical reliability cannot support digital inclusion. A technically strong system that ignores access relevance may still be unusable in classrooms or public offices. A system that lacks governance readiness may become unreliable as scripts, standards, and regional usage evolve. An equity-sensitive approach is especially important in low-resource NLP because the available corpus may reflect the most visible institutions rather than the full speech community. Dataset documentation is therefore necessary because data provenance, coverage, and known gaps shape downstream inclusion outcomes (Gebru et al.,2021).

5. Data Analysis: From Conversion Accuracy to Inclusion Friction

5.1 Reanalysis of Direction-Level Error Patterns

The benchmark results show a clear hierarchy across model families. Rule-based systems produce the highest average error burden because they cannot sufficiently model contextual phonology, morphological alternation, or irregular loanwords. Dictionary fusion systems improve some known lexical cases but remain limited when stems are affixed or when the input contains out-of-vocabulary terms. RNN models improve over rules and dictionaries by learning sequential patterns, while RNN-Attention models further reduce errors by allowing the decoder to focus on relevant source positions. The loanword-prompt neural conversion method performs best across all six conversion directions because it combines cross-attention with explicit lexical guidance. BLEU is not used as the main metric here because character and word errors are more transparent for script-conversion diagnosis (Papineni et al.,2002).

For a sociotechnical study, the important finding is not only that the best model is more accurate. The key point is that accuracy gains are unevenly meaningful across settings. Word-level errors are particularly important because a single wrong word can change a search query, a public-service instruction, or a student's understanding of a term. Character-level errors matter for readability and orthographic quality, but word-level errors better approximate task disruption. The following table therefore summarizes the best non-MCLP baseline for each direction and compares it with the prompt-guided model. Translation edit rate research is useful because it interprets errors through the number of human corrections needed to repair an output (Snover et al.,2006).

Table 3. Direction-level reanalysis of conversion performance

Pair	Best non-MCLP baseline	Baseline CER	Baseline WER	MCLP CER	MCLP WER	WER reduction
Ar->Cy	RNN-Attention	5.43	7.81	2.87	5.54	29.1%
Cy->Ar	RNN-Attention	4.68	6.08	2.50	4.33	28.8%
Ar->La	RNN-Attention	5.38	7.53	2.37	4.60	38.9%
La->Ar	RNN-Attention	4.16	5.39	2.02	3.72	31.0%
La->Cy	RNN-Attention	4.63	5.77	2.12	3.80	34.1%
Cy->La	RNN-Attention	3.25	5.28	1.82	3.16	40.2%

Table 3 indicates that the prompt-guided model improves every direction relative to the strongest non-MCLP baseline. The largest word-level reduction appears in the Cyrillic-to-Latin direction, where the best baseline WER is 5.28% and the prompt-guided model reduces it to 3.16%. The Arabic-to-Cyrillic direction remains the most difficult in absolute terms, with a prompt-guided WER of 5.54%, but even here the reduction compared with the best baseline is substantial. This pattern is consistent with the linguistic complexity of Arabic-based conversion, where vowel marking, regional lexical forms, and loanword normalization create additional ambiguity. Standardized reporting conventions are important because metric values are difficult to compare when tokenization and preprocessing differ (Post,2018).

The direction-level differences have direct inclusion implications. Arabic-to-Cyrillic and Arabic-to-Latin conversion are crucial for making Arabic-based regional materials searchable and usable in platforms that privilege Cyrillic or Latin input. Higher errors in these directions therefore imply greater friction for cultural archives, cross-border news access, and users educated in Arabic-based Kazakh. Conversely, relatively lower errors in Cyrillic-to-Latin conversion suggest that some modernization tasks may be easier to automate, although this should not be confused with social acceptance of a single Latin standard. Technical reliability can support reform, but it cannot decide which orthography a community should prefer. Meta-evaluation studies of machine translation also caution against treating automatic scores as sufficient evidence of real deployment quality (Marie et al.,2021).

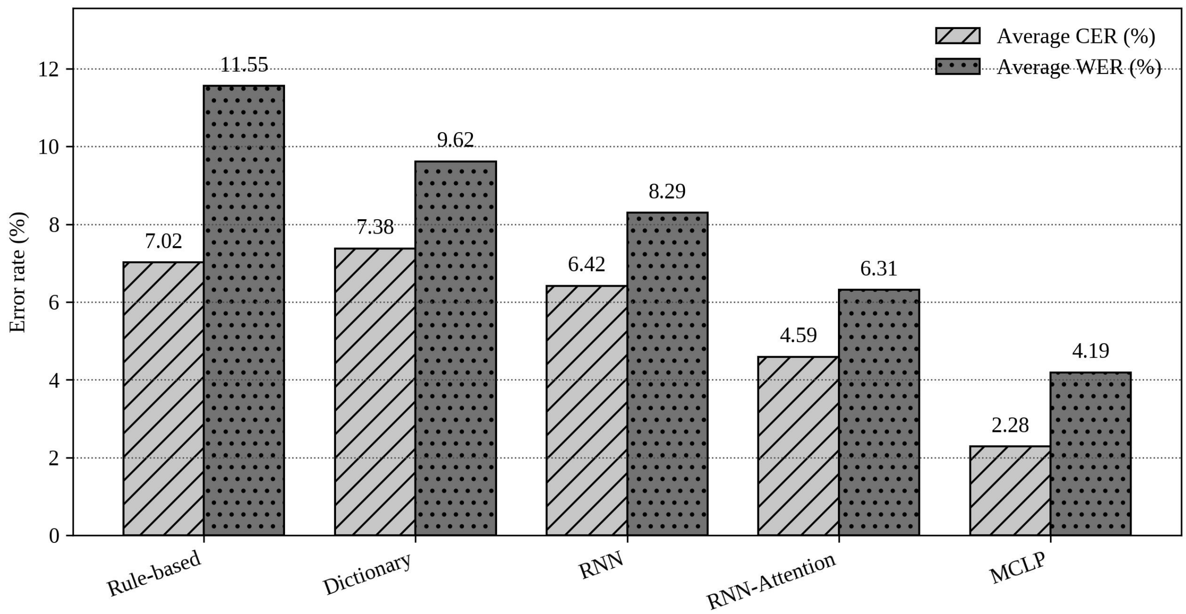

Figure 2. Average character and word error rates across six Kazakh conversion directions

Figure 2 aggregates the six directions and shows the overall reduction in error. The average WER declines from more than ten percent for rule-based conversion to approximately four percent for the prompt-guided Transformer model. In practical terms, this reduction changes the scale of user correction. In a 10,000-word corpus, the average word-level friction falls from more than 1,000 potentially wrong words under simple rule conversion to roughly 419 under the prompt-guided model. That is still not perfect, but it is a different usability regime. A teacher, archivist, or public officer may be able to review a few hundred uncertain tokens with targeted tools, whereas reviewing more than a thousand errors becomes a major labor burden. Large-scale human evaluation research further shows that context and expert judgment remain important even when automatic metrics improve (Freitag et al.,2021).

5.2 Script-Pair Asymmetry and Access Burden

Script conversion is not symmetric. The difficulty of converting from Arabic to Cyrillic is not identical to converting from Cyrillic to Arabic. This asymmetry matters because users often begin with different script positions. A Cyrillic-literate user accessing Arabic-based materials faces one set of conversion risks, while an Arabic-script user interacting with a Cyrillic-dominant public database faces another. Evaluation must therefore avoid reporting only a single average score. Inclusive design requires direction-aware reporting and user-facing warnings when certain directions are less reliable. Metric studies show that correlation with human judgment can vary by language pair, domain, and evaluation setting (Mathur et al.,2020).

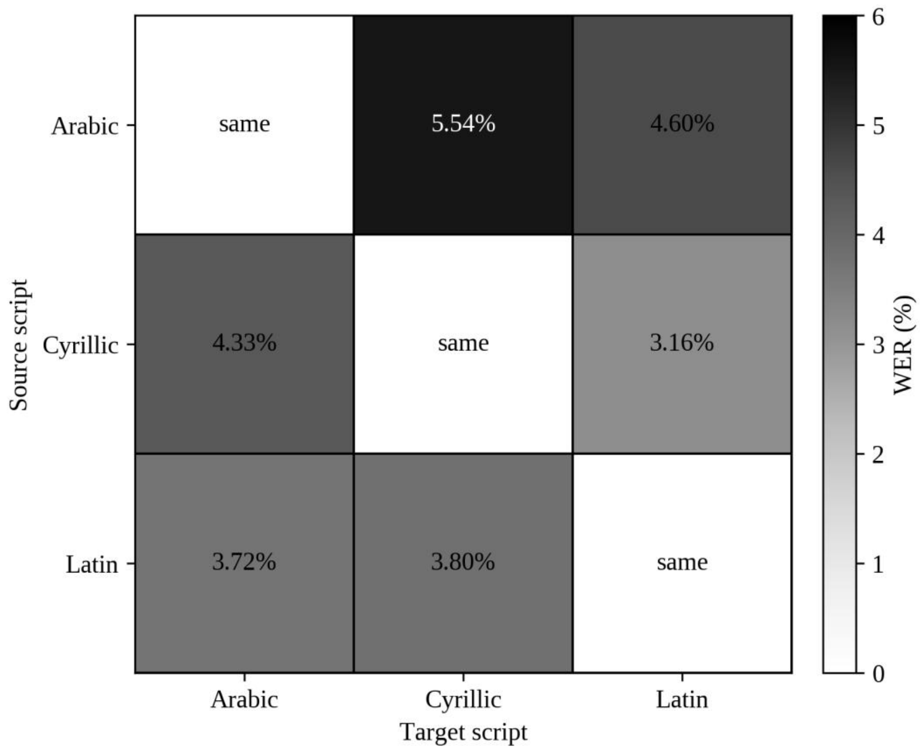

Figure 3. Direction-specific MCLP word error rates for Kazakh script conversion

Figure 3 visualizes the word error rates of the prompt-guided model by source and target script. The darker cells signal higher friction. The Arabic-to-Cyrillic direction has the highest word error rate among the six directions, followed by Arabic-to-Latin and Cyrillic-to-Arabic. The relatively light Cyrillic-to-Latin cell suggests lower friction for that direction. A digital inclusion policy should interpret these differences carefully. If an archive plans to convert Arabic-script materials into Cyrillic for a national library portal, it should budget for more human review and stronger metadata validation than a service converting Cyrillic classroom materials into Latin. Direction-specific risk reporting makes conversion infrastructure more honest and more usable. Neural evaluation frameworks are useful complements because they compare output quality beyond exact lexical matching (Rei et al.,2020).

A further implication concerns interface design. When conversion confidence varies by direction, a platform should not present all converted text with the same level of certainty. It may display low-confidence terms, loanwords, or proper names with review marks. It may allow users to choose regional variants. It may offer side-by-side comparison rather than replacing the source text entirely. These design choices are not cosmetic. They distribute responsibility between the automated system and the human user in a way that supports trust without overstating model certainty. Embedding-based generation metrics also support the idea that semantic adequacy should be evaluated alongside surface-level matching (Zhang et al.,2020).

5.3 Error Burden as Accessibility Friction

To make error rates interpretable for nontechnical stakeholders, this study converts average WER into a rough estimate of wrongly converted words per 10,000 words. The conversion is illustrative rather than predictive for every domain, because actual error distribution depends on text genre, name density, loanword frequency, and sentence length. Nevertheless, the estimate makes the scale of friction clearer. Error rates that appear small in a table may translate into hundreds of words requiring correction in a textbook chapter, legal archive, or digital library batch. Algorithmic auditing research provides a model for linking error analysis to accountability procedures (Raji et al.,2020).

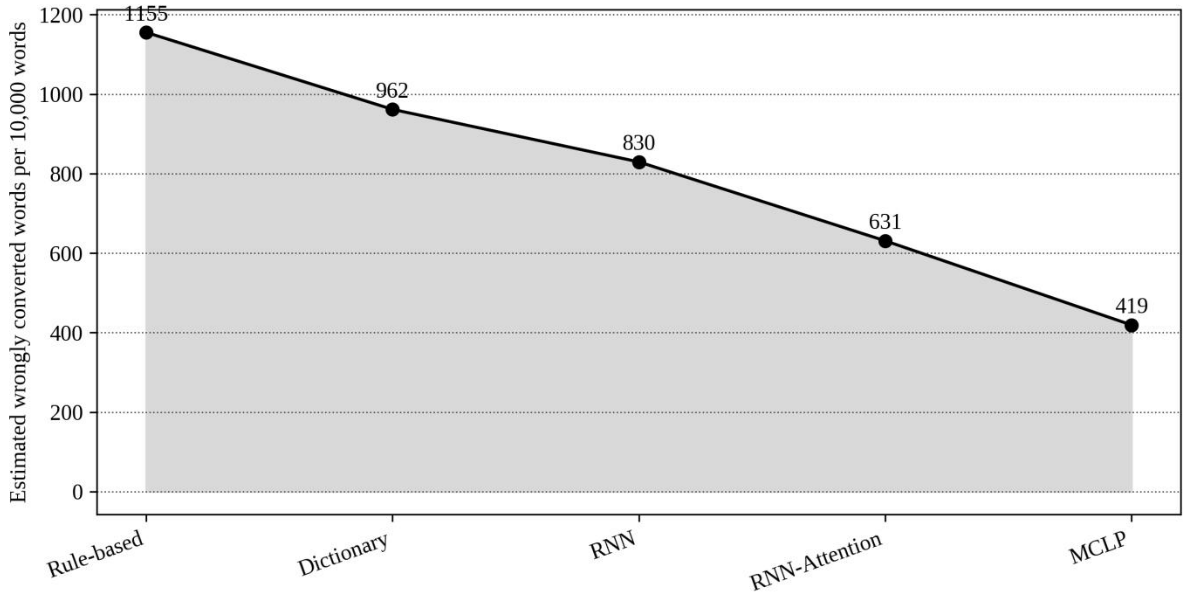

Figure 4. Estimated word-level conversion friction per 10,000 words

Figure 4 shows that the most advanced method reduces the review burden but does not eliminate it. This point is central to responsible deployment. A prompt-guided Transformer can make multiscript access far more feasible, yet it should be embedded in workflows that support review, correction, and feedback. In high-stakes settings, such as public records and educational examinations, automated conversion should be treated as a draft. In low-stakes discovery settings, such as search expansion or browsing, automated conversion may operate with less human intervention if uncertainty is clearly managed. The same system can therefore be deployed differently across risk levels. Industry-oriented fairness research indicates that practitioners need concrete tools, checklists, and escalation paths rather than abstract principles alone (Holstein et al.,2019).

Table 4. Inclusion-oriented interpretation of conversion errors

Metric signal	Technical meaning	Possible inclusion impact	Recommended control
High CER	Character-level mismatch or orthographic noise	Reduced readability and lower perceived quality	Spell-checking, script-aware OCR cleanup, visual comparison
High WER	Incorrect word output or wrong lexical choice	Search failure, wrong term, or incorrect administrative record	Loanword prompts, named-entity protection, human review
High direction asymmetry	One source-target path is less reliable	Unequal access for users of one script community	Direction-specific warnings and validation budgets
High loanword error	Borrowed or regional terms are unstable	Technical, medical, legal, or educational terms may be distorted	Domain lexicons and community-maintained alternatives
Low explainability	Output cannot be inspected or justified	Users cannot contest the conversion	Conversion logs, confidence display, and editable outputs

Table 4 translates common metric signals into governance and design controls. The table also shows why a sociotechnical framework does not reject technical metrics. It makes them more actionable. CER and WER remain important, but they must be connected to domain risk and user control. If a platform reports only aggregate accuracy, administrators may underestimate the labor needed to prepare reliable materials. If it reports direction-specific risk, loanword uncertainty, and named-entity alerts, users can decide when automation is sufficient and when expert review is necessary. Fairness research also warns that purely technical abstractions can hide social context and misstate who is harmed by system errors (Selbst et al.,2019).

6. Discussion: Where Script Conversion Produces Inclusion or Exclusion

6.1 Education and Learning Materials

See: https://inatgi.in/index.php/jtis/index for more information. https://doi.org/10.63646/jtis.2025.030303

Education is one of the most important deployment settings for Kazakh multiscript conversion. Students may encounter materials in one script at home, another script in school, and a third script in digital resources. Teachers may need to adapt content for mixed-script classrooms or cross-border learning communities. A reliable converter can reduce the cost of preparing parallel materials, especially for vocabulary lists, reading passages, cultural texts, and digital assignments. It can also support learners' transitioning between script systems by offering aligned text and highlighting differences. FLORES-style multilingual benchmarks illustrate the value of carefully curated evaluation sets for low-resource languages (Goyal et al.,2022).

However, educational use also exposes the limitations of automation. A converter that handles everyday words well may fail on subject terminology in science, history, mathematics, and information technology. Loanwords are common in academic vocabulary, and many terms have competing regional forms. If a converted text uses a nonstandard or unfamiliar term, students may learn inconsistent vocabulary. If the system silently changes to a proper name or historical concept, the error may persist in notes, assignments, and examinations. For this reason, educational deployment should include domain-specific lexicons, teacher review tools, and student-facing explanations of alternative spellings. Zero-shot multilingual translation research shows that shared models can improve access, but their behavior must be tested by direction and domain (Johnson et al.,2017).

Inclusive educational conversion should also consider cognitive load. A side-by-side interface can help learners compare scripts, but it can overwhelm younger users if too many variants are displayed. A staged design may be preferable: the main text appears in the selected script, difficult terms can be expanded to show equivalents, and loanwords can include brief notes when multiple accepted spellings exist. The goal is not only to deliver converted text but to support script literacy and language confidence. Large-scale human-centered translation research also suggests that language coverage should be treated as an inclusion goal, not simply a model-scaling target (Costa-jussa et al.,2022).

6.2 Public Services and Administrative Records

Public services require stricter reliability because conversion errors can affect names, addresses, certificates, benefits, and legal communication. Script differences may create duplicate records or mismatched identities if a person is registered in one script and searches or submits forms in another. A converter integrated into e-government systems must therefore protect named entities and preserve source-script metadata. Automatic conversion should never overwrite the original representation of a name without an auditable record. Low-resource evaluation datasets demonstrate why public-service conversion should be validated on named entities, forms, and domain-specific text rather than general prose alone (Guzman et al.,2019).

The risk is not only technical. A public service that accepts only one script may signal that other script users are less legitimate. Conversely, a service that offers conversion without explaining standards may generate distrust when citizens see unfamiliar forms of their names. Inclusive public-service design should allow users to view, verify, and correct converted forms. It should store script variants as linked fields, not as competing identities. Where official standards exist, the system should distinguish official transliteration from user-preferred display forms. This separation allows institutions to maintain consistency while respecting linguistic identity. Audits of web-crawled multilingual corpora show that low-resource language datasets can contain severe quality problems even when they appear large (Kreutzer et al.,2022).

Administrative conversion also requires accountability. Logs should record conversion engine version, lexicon version, confidence level, and human corrections. These logs allow agencies to diagnose recurring errors, update lexicons, and handle disputes. In this context, a prompt-guided Transformer is useful because its lexical prompt component can be maintained as a controlled resource. The lexicon becomes an institutional object that can be reviewed, versioned, and improved, rather than a hidden parameter inside a black-box model. Participatory machine-translation research demonstrates that community collaboration can improve both data quality and legitimacy (Nekoto et al.,2020). Blockchain governance studies offer a useful analogy for recording provenance, versioning, and accountability in distributed digital systems (Chen et al.,2024).

6.3 Cultural Heritage, Archives, and Search

Cultural heritage institutions face a different but equally important problem. Historical newspapers, magazines, manuscripts, oral-history transcripts, and family documents may use script conventions that differ from contemporary digital standards. Digitization projects often focus on scanning and OCR, but search access depends on cross-script metadata. If a user searches in Cyrillic but the archive stores Arabic-script titles without conversion, relevant materials may remain invisible. Conversely, if conversion is inaccurate, the archive may misrepresent names, places, and cultural terms. Research on language technologies for Indigenous languages shows that distance from dominant-resource settings requires careful design and documentation (Mager et al.,2018).

For archives, inclusive conversion should preserve the original script as the authoritative object while adding searchable converted layers. This layered approach respects the historical document and supports discovery. It also allows archivists to update conversion layers as models improve without altering the source record. Loanword-aware conversion is especially valuable in heritage collections because borrowed terms, personal names, and place names often carry historical traces. A purely standardized conversion may erase those traces, while a transparent system can display both normalized and original forms. Shared tasks for low-resource Indigenous languages also show the value of transparent benchmarks and community-specific test sets (Mager et al.,2021).

Platform searches introduce similar issues on a larger scale. Search engines and social platforms can use conversion to expand queries across scripts, improve recall, and connect communities. However, query expansion must be balanced against precision. If the system treats too many forms as equivalent, users receive irrelevant results. If it treats too few forms as equivalent, users miss relevant content. Direction-aware confidence scores and user-controlled script filters can make search more inclusive without flattening linguistic diversity. Open translation services demonstrate how public infrastructure can lower access barriers when tools remain transparent and reusable (Tiedemann and Thottingal,2020).

6.4 Platform Governance and Community Trust

Trust is a central condition for adopting language technology. Users are more likely to accept automated conversion when they can understand its purpose, inspect alternatives, and correct errors. Trust is weakened when a platform silently normalizes words, imposes one regional standard, or hides the original text. For Kazakh, where script choice is linked to history and identity, governance cannot be an afterthought. It should be designed into the system from the beginning. Critical work on large language models warns that scale alone cannot replace documentation, accountability, and community-centered evaluation (Bender et al.,2021). Information-systems research on blockchain governance similarly emphasizes auditability when digital services affect institutional trust (Lu,2022).

A community review model can improve both accuracy and legitimacy. Linguists, teachers, editors, archivists, software developers, and script-community representatives can jointly review loanword dictionaries, named-entity lists, and domain terminology. Corrections submitted by users should be moderated and converted into lexicon updates only after quality control. The system should document which sources support each lexical entry, when it was updated, and which regional or domain labels apply. This kind of governance turns conversion from a one-time engineering project into a living language infrastructure. Decolonial language-technology research further stresses that affected communities should help define the goals and limits of language tools (Bird,2020).

The governance challenge is especially important for low-resource NLP. Large language models may appear attractive because they can perform conversion-like tasks with minimal engineering. Yet they may hallucinate, ignore local standards, or produce inconsistent outputs unless constrained by high-quality resources. A responsible multiscript system should therefore combine neural capability with structured lexicons, corpus documentation, and human oversight. The point is not to reject advanced models, but to place them within accountable language stewardship. Bias research in NLP shows that harms must be defined with reference to affected communities rather than assumed from aggregate metrics (Blodgett et al.,2020). Multilingual representation studies also show that nominal language coverage does not guarantee equal model quality across languages (Wu and Dredze,2020). FinTech scholarship offers a parallel reminder that digital innovation can increase access only when governance and user protection are designed together (Kou and Lu,2025).

Table 5. Deployment recommendations by use setting

Use setting	Primary inclusion goal	Main risk	Recommended deployment model
Education	Parallel learning materials and script literacy	Incorrect terminology and student confusion	Teacher-reviewed conversion with domain glossaries
Public services	Reliable access to forms, names, and notices	Identity mismatch and legal ambiguity	Original-script preservation with auditable official variants
Cultural archives	Discoverability of historical and regional materials	Loss of source-script meaning and metadata gaps	Layered metadata with searchable converted fields
Platform search	Cross-script retrieval and community connection	Low precision or missed results	Query expansion with confidence scores and filters
Mobile communication	Everyday readability across script users	Overcorrection of informal language	User-selectable profiles and editable output

Table 5 emphasizes that deployment should be differentiated. A single converter may serve multiple domains, but the interface, review threshold, and governance controls should change according to risk. For education, the system should support learning and explanation. For public services, it should prioritize identity reliability. For archives, it should preserve source-script authenticity. For search, it should balance recall and precision. For mobile communication, it should respect user preference and allow rapid editing. Inclusive technology is therefore not a universal setting; it is an adaptive arrangement of model, data, interface, and institution. Surveys of multilingual NMT indicate that shared models, transfer learning, and domain adaptation must be balanced against language-specific quality control (Dabre et al.,2020).

7. Design Principles for Inclusive Kazakh Multiscript Technologies

The analysis above suggests six design principles. The first is source preservation. Converted text should not replace the source script in important records. Instead, systems should maintain aligned source and target forms so that users can verify conversion and institutions can audit changes. Source preservation is especially important for archives, public records, and research corpora. Massive multilingual NMT research shows that a single model can support many directions, but deployment still requires direction-specific monitoring (Aharoni et al.,2019).

The second principle is direction-aware confidence. Since conversion difficulty differs by source and target script, systems should report uncertainty by direction. A platform should not imply that Arabic-to-Cyrillic conversion is equally reliable as Cyrillic-to-Latin conversion if benchmark evidence shows otherwise. Direction-aware confidence can guide human review, interface warnings, and resource allocation. Unsupervised translation studies are relevant because they show how cross-lingual structure can be exploited when direct parallel data are limited (Lample et al.,2018).

The third principle is loanword transparency. Loanwords should not be treated as ordinary tokens when they are frequent sources of ambiguity. A converter should identify loanword matches, display standardized alternatives where needed, and record the lexicon source. This principal links technical prompt construction with governance because the loanword dictionary becomes a public-facing resource rather than a hidden component. Zero-resource translation research is useful for Kazakh conversion because some script pairs and domain combinations have limited aligned data (Firat et al.,2016).

The fourth principle is named-entity protection. Names of people, places, institutions, and cultural objects are central to identity and search. Errors in named entities have higher inclusion costs than errors in many common words. Conversion systems should therefore include named-entity recognition, protected name lists, manual review options, and metadata fields that preserve alternative spellings. Transfer learning remains important when low-resource conversion models must benefit from related high-resource tasks without erasing local orthographic features (Zoph et al.,2016).

The fifth principle is community-correctable infrastructure. Users should be able to report errors, but their corrections should be curated rather than immediately absorbed. A correction workflow should identify the affected script pair, domain, token type, and context. Expert review can then decide whether the correction reflects a true error, an acceptable regional variant, or a domain-specific preference. This process creates an evolving knowledge base for the language community. Neural machine translation research also identifies data sparsity, rare words, and domain shift as central challenges for low-resource tasks (Koehn and Knowles,2017).

The sixth principle is accessibility beyond the model. A highly accurate converter is not inclusive if it requires constant internet access, supports only one platform, lacks keyboard integration, or renders fonts poorly. Inclusive deployment should include offline modes where possible, Unicode-compliant font handling, mobile-friendly design, keyboard support, and APIs that allow schools, libraries, and local developers to integrate conversion into their own services. Open statistical MT toolkits remain relevant because they support transparent baselines and reproducible comparisons (Koehn et al.,2007). IoT security research provides a parallel case in which technical interoperability must be paired with governance, documentation, and risk control (Xu et al.,2021).

Table 6. Governance checklist for accountable multiscript conversion

Checklist item	Implementation detail	Expected benefit
Corpus documentation	Describe sources, genres, scripts, regions, and update cycles	Clarifies coverage and limits
Lexicon versioning	Record loanword and named-entity dictionary versions	Supports audit and reproducibility
Error typology	Separate vowel, consonant, loanword, name, and segmentation errors	Guides targeted improvement
User correction channel	Allow users to submit corrections with context	Improve legitimacy and local fit
Domain profiles	Maintain education, public service, archive, and search settings	Matches risk level to deployment
Human review thresholds	Trigger review for low confidence or high-stakes categories	Prevents silent harm
Original text retention	Store source script alongside converted variants	Protects identity and cultural authenticity

Table 6 provides a practical checklist for institutions that intend to deploy Kazakh conversion services. The checklist is intentionally operational. It can be used by a university creating parallel learning resources, a public office building a bilingual portal, an archive digitizing regional newspapers, or a software company building search tools. The checklist also encourages responsibility sharing. Model developers are not the only actors who shape inclusion; editors, data stewards, administrators, teachers, and users all participate in maintaining the conversion infrastructure. Fast open-source sequence-modeling toolkits can support reproducible experiments and practical deployment pipelines for language technologies (Junczys-Dowmunt et al.,2018). Audit technology studies also reinforce the value of traceability and verification when digital systems are deployed in organizations (Wu et al.,2025).

8. Limitations and Future Research

This study has several limitations. First, it relies on secondary benchmark data rather than collecting new user-experience data. The analysis shows how technical error rates can be interpreted as inclusion friction, but it does not measure how real users respond to specific conversion errors in classrooms, public offices, archives, or search interfaces. Future work should conduct user studies with readers from different script backgrounds to measure readability, trust, correction behavior, and perceived legitimacy. Sequence-modeling toolkits lower implementation barriers, but they do not remove the need for careful data design and human review (Ott et al.,2019).

Second, the analysis treats word error rate as a useful proxy for access friction, but not all word errors are equal. A wrong suffix, a wrong loanword, and a wrong personal name have different consequences. Future evaluation should weight errors by domain risk, token type, and user task. A sociotechnical benchmark for Kazakh conversion could include named entities, school terminology, health information, administrative phrases, historical documents, and informal communication. Such a benchmark would make model comparison more relevant to inclusion. Transformer software ecosystems have made multilingual experimentation easier, yet ease of use can also obscure data and evaluation weaknesses (Wolf et al.,2020).

Third, the paper does not resolve the normative question of which Latin standard or regional spelling should be treated as canonical. That question belongs to language communities, educators, institutions, and policy processes. The technical contribution is to show how systems can preserve alternatives, display uncertainty, and support transparent governance. Future research should examine how communities negotiate competing standards and how software can support those negotiations without imposing a hidden hierarchy. Studies of multilingual representation learning show that cross-lingual models can transfer knowledge but still underperform for underrepresented languages (Pires et al.,2019).

Fourth, more work is needed on multimodal pipelines. Many heritage materials require OCR before conversion, and OCR errors interact with script conversion errors. Speech technologies also matter because script choice affects subtitles, voice assistants, and transcription tools. Future multiscript inclusion research should connect OCR, speech recognition, machine translation, search, and script conversion into a unified language infrastructure perspective. Cross-lingual embedding surveys further show that alignment quality depends on language relatedness, data availability, and evaluation choices (Ruder et al.,2019).

9. Conclusion

Kazakh multiscript conversion is often framed as an engineering task given an input in Arabic, Cyrillic, or Latin script, generate the corresponding output in another script. This paper has argued that the task is also a digital inclusion problem. Script conversion determines whether people can search, read, learn, submit forms, preserve cultural memory, and maintain continuity across script communities. Technical accuracy is therefore necessary but not sufficient. A converter becomes inclusive only when it is embedded in transparent, accountable, and community-sensitive infrastructure. Energy and policy studies of NLP also remind developers that language technology has environmental and institutional costs.

The secondary analysis of benchmark results shows that loanword-prompt neural conversion substantially reduces both character-level and word-level errors across six conversion directions. This improvement is important because it lowers the practical correction burden and makes cross-script services more feasible. However, the remaining direction-level asymmetries demonstrate that deployment should be risk-aware. Arabic-based source materials, loanwords, regional spellings, and named entities require particular attention. Automated conversion should be paired with source preservation, confidence reporting, lexicon governance, and human review in high-stakes settings. Digital-divide research shows that access gains are meaningful only when they translate into actual participation and outcomes.

The broader contribution of the paper is a sociotechnical framework linking script ecology, linguistic resources, conversion services, access settings, governance, and inclusion outcomes. This framework can guide future research and institutional deployment. For Kazakh and other multiscript languages, digital inclusion will not be achieved by a model alone. It will depend on the long-term stewardship of corpora, standards, interfaces, and communities. Multiscript language technologies should therefore be treated as digital public infrastructure for linguistic equity. AI survey research supports the conclusion that language technologies should be evaluated as evolving systems embedded in applications and institutions.

Acknowledgement

The authors thank the editors and anonymous reviewers for their constructive comments. The authors also acknowledge the broader research community working on Kazakh language technologies, low-resource NLP, and digital inclusion. No external funding was received for this study. The authors also recognize that language technologies depend on long-term interdisciplinary work across artificial intelligence, information systems, and society.

Data Availability Statement

All derived analytical tables and figures used in this article are included in the manuscript. The study uses secondary benchmark values reported in the motivating Kazakh multiscript conversion manuscript and does not involve newly collected personal data. Future releases of evaluation data should therefore include documentation on collection, cleaning, and access controls for sensitive script-conversion examples.

Conflict of Interest

The authors declare that they have no conflict of interest.

Reference

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. Information Systems Research, 7(1), 111-134. https://doi.org/10.1287/isre.7.1.111

Kornai, A. (2013). Digital language death. PLOS ONE, 8(10), e77056. https://doi.org/10.1371/journal.pone.0077056

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1409.0473

Dong, J., Jiang, T., Cheng, L., Anwar, A., & Yang, Y. (2018). A compromise Arabic-Kazakh coded character processing method based on the OpenType font format. Computer Standards & Interfaces, 55, 1-7. https://doi.org/10.1016/j.csi.2017.02.005

Karimi, S., Scholer, F., & Turpin, A. (2011). Machine transliteration survey. ACM Computing Surveys, 43(3), Article 17. https://doi.org/10.1145/1922649.1922654

Al-Onaizan, Y., & Knight, K. (2002). Machine transliteration of names in Arabic texts. Proceedings of the ACL-02 Workshop on Computational Approaches to Semitic Languages. https://doi.org/10.3115/1118637.1118642

Akhmed-Zaki, D., Mansurova, M., Pyrkova, A., & Yarullin, R. (2021). Development of the information system for the Kazakh language corpus. Cogent Engineering, 8(1), 1896418. https://doi.org/10.1080/23311916.2021.1896418

Bapna, A., & Firat, O. (2019). Simple, scalable adaptation for neural machine translation. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 1538-1548. https://doi.org/10.18653/v1/D19-1165

Kumaran, A., & Kellner, T. (2007). A generic framework for machine transliteration. Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 721-722. https://doi.org/10.1145/1277741.1277876

Knight, K., & Graehl, J. (1998). Machine transliteration. Computational Linguistics, 24(4), 599-612. https://doi.org/10.5555/972764.972767

Sennrich, R., Haddow, B., & Birch, A. (2016). Neural machine translation of rare words with subword units. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 1715-1725. https://doi.org/10.18653/v1/P16-1162

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 1724-1734. https://doi.org/10.3115/v1/D14-1179

Hargittai, E. (2002). Second-level digital divide: Differences in people's online skills. First Monday, 7(4). https://doi.org/10.5210/fm.v7i4.942

Selwyn, N. (2004). Reconsidering political and popular understandings of the digital divide. New Media & Society, 6(3), 341-362. https://doi.org/10.1177/1461444804042519

Helsper, E. J. (2012). A corresponding fields model for the links between social and digital exclusion. Communication Theory, 22(4), 403-426. https://doi.org/10.1111/j.1468-2885.2012.01416.x

van Deursen, A. J. A. M., & van Dijk, J. A. G. M. (2014). The digital divide shifts to differences in usage. New Media & Society, 16(3), 507-526. https://doi.org/10.1177/1461444813487959

Scheerder, A., van Deursen, A., & van Dijk, J. (2017). Determinants of internet skills, uses and outcomes: A systematic review of the second- and third-level digital divide. Telematics and Informatics, 34(8), 1607-1624. https://doi.org/10.1016/j.tele.2017.07.007

Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. Organization Science, 3(3), 398-427. https://doi.org/10.1287/orsc.3.3.398

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 6282-6293. https://doi.org/10.18653/v1/2020.acl-main.560

Ranathunga, S., Lee, E. S. A., Skenduli, M. P., Shekhar, R., Alam, M., & Kaur, R. (2023). Neural machine translation for low-resource languages: A survey. ACM Computing Surveys, 55(11), Article 229. https://doi.org/10.1145/3567592

Haddow, B., Birch, A., Bawden, R., Bogoychev, N., Currey, A., Germann, U., Grundkiewicz, R., Heafield, K., Miceli Barone, A. V., & Williams, P. (2022). Survey of low-resource machine translation. Computational Linguistics, 48(3), 673-732. https://doi.org/10.1162/coli_a_00446

Dror, R., Baumer, G., Bogomolov, M., & Reichart, R. (2018). The hitchhiker's guide to testing statistical significance in natural language processing. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 1383-1392. https://doi.org/10.18653/v1/P18-1128

Galla, C. K. (2016). Indigenous language revitalization, promotion, and education: Function of digital technology. Computer Assisted Language Learning, 29(7), 1137-1151. https://doi.org/10.1080/09588221.2016.1166137

Toiganbayeva, N., Makhambetov, O., Seisenbekov, Z., & Varol, H. A. (2022). KOHTD: Kazakh offline handwritten text dataset. Image and Vision Computing, 123, 116827. https://doi.org/10.1016/j.image.2022.116827

Yeshpanov, R., Polonskaya, A., & Varol, H. A. (2024). KazParC: Kazakh parallel corpus for machine translation. Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, 9633-9644. https://doi.org/10.18653/v1/2024.lrec-main.842

Veitsman, Y., & Hartmann, M. (2025). Recent advancements and challenges of Turkic Central Asian language processing. Proceedings of the First Workshop on Language Models for Low-Resource Languages, 295-307. https://doi.org/10.18653/v1/2025.loreslm-1.25

Senel, L. K., Ebing, B., Baghirova, K., Schuetze, H., & Glavas, G. (2024). Kardes-NLU: Transfer to low-resource languages with the help of a high-resource cousin. Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics, 1776-1790. https://doi.org/10.18653/v1/2024.eacl-long.100

Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. Organization Science, 11(4), 404-428. https://doi.org/10.1287/orsc.11.4.404.14600

Geels, F. W. (2004). From sectoral systems of innovation to socio-technical systems: Insights about dynamics and change from sociology and institutional theory. Research Policy, 33(6-7), 897-920. https://doi.org/10.1016/j.respol.2004.01.015

Markard, J., Raven, R., & Truffer, B. (2012). Sustainability transitions: An emerging field of research and its prospects. Research Policy, 41(6), 955-967. https://doi.org/10.1016/j.respol.2012.02.013

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency, 220-229. https://doi.org/10.1145/3287560.3287596

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daume III, H., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92. https://doi.org/10.1145/3458723

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: A method for automatic evaluation of machine translation. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, 311-318. https://doi.org/10.3115/1073083.1073135

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. Proceedings of AMTA 2006, 223-231. https://doi.org/10.3115/1626355.1626389

Post, M. (2018). A call for clarity in reporting BLEU scores. Proceedings of the Third Conference on Machine Translation, 186-191. https://doi.org/10.18653/v1/W18-6319

Marie, B., Fujita, A., & Rubino, R. (2021). Scientific credibility of machine translation research: A meta-evaluation of 769 papers. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics, 7297-7306. https://doi.org/10.18653/v1/2021.acl-long.566

Freitag, M., Foster, G., Grangier, D., Ratnakar, V., Tan, Q., & Macherey, W. (2021). Experts, errors, and context: A large-scale study of human evaluation for machine translation. Transactions of the Association for Computational Linguistics, 9, 1460-1474. https://doi.org/10.1162/tacl_a_00437

Mathur, N., Baldwin, T., & Cohn, T. (2020). Tangled up in BLEU: Reevaluating the evaluation of automatic machine translation evaluation metrics. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 4984-4997. https://doi.org/10.18653/v1/2020.acl-main.448

Rei, R., Stewart, C., Farinha, A. C., & Lavie, A. (2020). COMET: A neural framework for MT evaluation. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 2685-2702. https://doi.org/10.18653/v1/2020.emnlp-main.213

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2020). BERTScore: Evaluating text generation with BERT. International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1904.09675

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 33-44. https://doi.org/10.1145/3351095.3372873

Holstein, K., Wortman Vaughan, J., Daume III, H., Dudik, M., & Wallach, H. (2019). Improving fairness in machine learning systems: What do industry practitioners need? Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1-16. https://doi.org/10.1145/3290605.3300830

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. Proceedings of the Conference on Fairness, Accountability, and Transparency, 59-68. https://doi.org/10.1145/3287560.3287598

Goyal, N., Gao, C., Chaudhary, V., Chen, P. J., Wenzek, G., Ju, D., Krishnan, S., Ranzato, M., Guzman, F., & Fan, A. (2022). The FLORES-101 evaluation benchmark for low-resource and multilingual machine translation. Transactions of the Association for Computational Linguistics, 10, 522-538. https://doi.org/10.1162/tacl_a_00474

Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viegas, F., Wattenberg, M., Corrado, G., Hughes, M., & Dean, J. (2017). Google's multilingual neural machine translation system: Enabling zero-shot translation. Transactions of the Association for Computational Linguistics, 5, 339-351. https://doi.org/10.1162/tacl_a_00065

Costa-jussa, M. R., Cross, J., Celebi, O., Elbayad, M., Heafield, K., Heffernan, K., Kalbassi, E., Lam, J., Licht, D., Maillard, J., Sun, A., Wang, S., Wenzek, G., Youngblood, A., Akula, B., Barrault, L., Mejia-Gonzalez, G., Hansanti, P., Hoffman, J., Jarrett, S., et al. (2022). No language left behind: Scaling human-centered machine translation. Transactions of the Association for Computational Linguistics, 10, 1339-1359. https://doi.org/10.1162/tacl_a_00559

Guzman, F., Chen, P. J., Ott, M., Pino, J., Lample, G., Koehn, P., Chaudhary, V., & Ranzato, M. (2019). The FLORES evaluation datasets for low-resource machine translation: Nepali-English and Sinhala-English. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, 6098-6111. https://doi.org/10.18653/v1/D19-1632

Kreutzer, J., Caswell, I., Wang, L., Wahab, A., van Esch, D., Ulzii-Orshikh, N., Tapo, A., Subramani, N., Sokolov, A., Sikasote, C., Setyawan, M., Sarin, S., Samb, S., Sagot, B., Rivera, C., Rios, A., Papadimitriou, I., Osei, S., Ortiz Suarez, P., et al. (2022). Quality at a glance: An audit of web-crawled multilingual datasets. Transactions of the Association for Computational Linguistics, 10, 50-72. https://doi.org/10.1162/tacl_a_00447

Nekoto, W., Marivate, V., Matsila, T., Fasubaa, T., Kolawole, T., Fagbohungbe, T., Akinola, S. O., Muhammad, S. H., Kabongo, S., Osei, S., Freshia, S., Niyongabo, R. A., Macharm, R., Ogayo, P., Ahia, O., et al. (2020). Participatory research for low-resourced machine translation: A case study in African languages. Findings of the Association for Computational Linguistics: EMNLP 2020, 2144-2160.

https://doi.org/10.18653/v1/2020.findings-emnlp.195

Mager, M., Gutierrez-Vasques, X., Sierra, G., & Meza, I. (2018). Challenges of language technologies for the indigenous languages of the Americas. Proceedings of the 27th International Conference on Computational Linguistics, 55-69. https://doi.org/10.18653/v1/C18-1006

Tiedemann, J., & Thottingal, S. (2020). OPUS-MT: Building open translation services for the world. Proceedings of the 22nd Annual Conference of the European Association for Machine Translation, 479-480. https://doi.org/10.18653/v1/2020.eamt-1.61

Mager, M., Oncevay, A., Ebrahimi, A., Ortega, J., Rios, A., Fan, A., Gutierrez-Vasques, X., Chiruzzo, L., Gimenez-Lugo, G., Ramos, R., et al. (2021). Findings of the AmericasNLP 2021 shared task on open machine translation for indigenous languages of the Americas. Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas, 202-217. https://doi.org/10.18653/v1/2021.americasnlp-1.23

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610-623. https://doi.org/10.1145/3442188.3445922

Bird, S. (2020). Decolonising speech and language technology. Proceedings of the 28th International Conference on Computational Linguistics, 3504-3519. https://doi.org/10.18653/v1/2020.coling-main.313

Blodgett, S. L., Barocas, S., Daume III, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of bias in NLP. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 5454-5476. https://doi.org/10.18653/v1/2020.acl-main.485

Wu, S., & Dredze, M. (2020). Are all languages created equal in multilingual BERT? Proceedings of the 5th Workshop on Representation Learning for NLP, 120-130. https://doi.org/10.18653/v1/2020.repl4nlp-1.16

Dabre, R., Chu, C., & Kunchukuttan, A. (2020). A survey of multilingual neural machine translation. ACM Computing Surveys, 53(5), Article 99. https://doi.org/10.1145/3406095

Aharoni, R., Johnson, M., & Firat, O. (2019). Massively multilingual neural machine translation. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, 3874-3884. https://doi.org/10.18653/v1/N19-1388

Lample, G., Ott, M., Conneau, A., Denoyer, L., & Ranzato, M. (2018). Phrase-based and neural unsupervised machine translation. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 5039-5049. https://doi.org/10.18653/v1/D18-1549

Firat, O., Sankaran, B., Al-Onaizan, Y., Vural, F. T. Y., & Cho, K. (2016). Zero-resource translation with multi-lingual neural machine translation. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 268-277. https://doi.org/10.18653/v1/D16-1026

Zoph, B., Yuret, D., May, J., & Knight, K. (2016). Transfer learning for low-resource neural machine translation. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 1568-1575. https://doi.org/10.18653/v1/D16-1163

Koehn, P., & Knowles, R. (2017). Six challenges for neural machine translation. Proceedings of the First Workshop on Neural Machine Translation, 28-39. https://doi.org/10.18653/v1/W17-3204

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., & Herbst, E. (2007). Moses: Open source toolkit for statistical machine translation. Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume, 177-180. https://doi.org/10.3115/1557769.1557821

Junczys-Dowmunt, M., Grundkiewicz, R., Dwojak, T., Hoang, H., Heafield, K., Neckermann, T., Seide, F., Germann, U., Fikri Aji, A., Bogoychev, N., Martins, A. F. T., & Birch, A. (2018). Marian: Fast neural machine translation in C++. Proceedings of ACL 2018, System Demonstrations, 116-121. https://doi.org/10.18653/v1/P18-4020

Ott, M., Edunov, S., Baevski, A., Fan, A., Gross, S., Ng, N., Grangier, D., & Auli, M. (2019). fairseq: A fast, extensible toolkit for sequence modeling. Proceedings of NAACL-HLT 2019: Demonstrations, 48-53. https://doi.org/10.18653/v1/N19-4009

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., et al. (2020). Transformers: State-of-the-art natural language processing. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 38-45. https://doi.org/10.18653/v1/2020.emnlp-demos.6

Pires, T., Schlinger, E., & Garrette, D. (2019). How multilingual is multilingual BERT? Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 4996-5001. https://doi.org/10.18653/v1/P19-1493

Ruder, S., Vulic, I., & Sogaard, A. (2019). A survey of cross-lingual word embedding models. Journal of Artificial Intelligence Research, 65, 569-631. https://doi.org/10.1613/jair.1.11640

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3645-3650. https://doi.org/10.18653/v1/P19-1355

van Dijk, J. A. G. M. (2006). Digital divide research, achievements and shortcomings. Poetics, 34(4-5), 221-235. https://doi.org/10.1016/j.poetic.2006.05.004

Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1-29. https://doi.org/10.1080/23270012.2019.1570365

Chen, Y., Lu, Y., Bulysheva, L., & Kataev, M. Y. (2024). Applications of blockchain in Industry 4.0: A review. Information Systems Frontiers, 26(5), 1715-1729. https://doi.org/10.1007/s10796-022-10248-7

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224. https://doi.org/10.1016/j.jii.2021.100224

Kou, G., & Lu, Y. (2025). FinTech: A literature review of emerging financial technologies and applications. Financial Innovation, 11, Article 1. https://doi.org/10.1186/s40854-024-00668-6

Xu, L. D., Lu, Y., & Li, L. (2021). Embedding blockchain technology into IoT for security: A survey. IEEE Internet of Things Journal, 8(13), 10452-10473. https://doi.org/10.1109/JIOT.2021.3060508