

The Azerbaijani Language in Artificial Intelligence: Challenges, Opportunities and Strategic Implications

Yedgar. M. Jafarov

Institute of Linguistics named after Nasimi, Azerbaijan National Academy of Sciences

ABSTRACT: This study examines the role of the Azerbaijani language in the context of artificial intelligence (AI), with a particular focus on natural language processing (NLP), linguistic resources, and issues of technological sovereignty. The research highlights that while AI technologies are rapidly transforming global communication and decision-making systems, low-resource languages such as Azerbaijani remain underrepresented in digital infrastructures. The paper analyzes the structural, morphological and semantic features of the Azerbaijani language, emphasizing its agglutinative nature and its potential as a bridging language within the Turkic language family.

In addition, the study identifies key challenges, including the lack of national corpora, annotated datasets, thesauri, frequency dictionaries and domain-specific linguistic models, which limit the effective integration of Azerbaijani into AI systems. Particular attention is given to the risks associated with reliance on foreign AI infrastructures, including semantic misalignment, decision-making inaccuracies and potential threats to national sovereignty. The concept of digital colonialism is explored as a broader framework for understanding these risks.

Furthermore, the paper argues that the development of AI systems in the Azerbaijani language is essential not only for technological advancement but also for preserving cultural identity, ensuring accurate governance, and enabling context-sensitive decision-making. It proposes that investment in national language technologies, including text analysis tools, speech processing systems and locally trained AI models, is a strategic necessity. The findings suggest that strengthening the digital presence of Azerbaijani will contribute to the formation of independent technological ecosystems and enhance the country's participation in the global digital economy.

KEYWORDS: Azerbaijani language; Artificial Intelligence; Natural Language Processing; low-resource languages; digital sovereignty; digital colonialism; text analysis; linguistic resources; Turkic languages; AI governance

Artificial intelligence technologies have, over recent decades, advanced at an unprecedented pace and have become an integral component of contemporary everyday life. Given that language constitutes the primary medium of human communication, language-based applications of artificial intelligence assume particular significance. At the core of these applications lies Natural Language Processing (NLP). Currently widely recognized under this abbreviation, NLP represents a multidisciplinary subfield situated at the intersection of computer science, artificial intelligence and linguistics Natural Language Processing [5, s. 3].

NLP may be defined as the computational analysis of the structural properties and established rules of natural languages, as well as their subsequent adaptation for technical purposes. Solutions developed on the basis of such analyses enable a wide range of applications, including automatic machine translation, the development of question-answering systems, speech recognition and command interpretation, speech synthesis, text summarization and the protection and processing of information.

The extensive use of information technologies and social media platforms has significantly increased reliance on written communication, more than at any previous point in history. This trend has facilitated the integration of tools such as automated correction, editing systems, and speech-to-text applications into digital communication environments. Such technologies, grounded in NLP, contribute to the analysis of written text and the verification of linguistic norms. In particular, speech recognition and command interpretation systems are expected, in the foreseeable future, to reduce or even eliminate the need for traditional input devices such as keyboards and mice. This transformation is likely to bring about profound changes in human-computer interaction, making computational systems more accessible to a broader population.

At present, NLP is applied across a wide array of domains, including sentiment analysis, chatbots, speech recognition, machine translation, pronunciation assessment, search engine optimization (SEO) keyword identification, information extraction, and advertisement relevance optimization [6, s. 61–63].

The Azerbaijani Language in Artificial Intelligence: Challenges, Opportunities and Strategic Implications

From a structural perspective, NLP encompasses two principal components: (1) natural language understanding and (2) natural language generation. Core processes within these components include tokenization, stemming, lemmatization, grammatical and contextual disambiguation, named entity recognition, and syntactic composition. The effective processing of natural language necessitates sequential modeling across multiple linguistic levels phonetic, phonological, morphological, syntactic and semantic as well as the development of robust system architectures. In this regard, it is imperative that comprehensive projects addressing these dimensions be implemented for the Azerbaijani language.

Like many other national languages, the Azerbaijani language occupies a relatively modest position in the technical development of artificial intelligence, while simultaneously holding substantial strategic and cultural importance. This duality is primarily associated with the issue of linguistic representation in digital environments. A foundational principle in NLP development posits that the extent to which a language is represented in digital corpora directly influences the degree to which artificial intelligence systems can effectively “learn” and process that language. In this context, Azerbaijani is classified among low-resource languages. This classification stems from the absence of large-scale linguistic corpora encompassing billions of words, annotated datasets, as well as extensive speech and dialogue repositories. Consequently, languages such as Azerbaijani remain significantly disadvantaged in comparison to global languages such as English, Russian, and Chinese in the competitive landscape of artificial intelligence development.

The development of the Azerbaijani language within the domain of artificial intelligence constitutes not merely a technological concern but, to a significant extent, an issue of sovereignty. In the near future, it is highly plausible that a wide range of public services, as well as legal relations among relevant entities, will be managed through artificial intelligence systems. As an active participant in global development processes, Azerbaijan is likewise progressing in this direction. In this context, the Azerbaijani language serves as a foundational medium for the transfer of national legal and administrative frameworks along with the corresponding regulatory terminology into artificial intelligence systems [8].

In the contemporary era, the integration of state languages into technological infrastructures represents one of the most effective mechanisms for mitigating the risks of digital colonialism. It should be noted that Digital Colonialism refers to the economic, cultural, and political influence exerted by technologically advanced nations over less developed countries through digital tools and infrastructures [3, s.211].

Such influence is largely facilitated through the control and governance of digital resources within these countries. Major technology corporations consolidate their power by collecting and exploiting user data from less developed regions. This asymmetry in data ownership gradually begins to shape economic and political trajectories at both national and individual levels. The transfer of such data predominantly to developed countries contributes to monopolistic tendencies in local markets, thereby undermining domestic economic structures.

Digital colonialism is further associated with the increasing dominance of global technology companies in international markets. By penetrating and, in some cases, overtaking the economies of developing countries, these corporations establish control over local services and products. For instance, global social media platforms and information retrieval systems frequently dominate domestic markets, which may result in the erosion of local employment opportunities and economic autonomy.

Over time, digital colonialism may also lead to the marginalization of local languages and cultural identities within digital ecosystems. This phenomenon is often linked to the dominance of global platforms such as Facebook, Google, and Amazon, whose widespread influence can result in the underrepresentation of local languages, traditions, and cultural practices in digital environments.

This global challenge can thus be conceptualized as technological colonialism, wherein the economic and social influence of developed nations exerted through digital infrastructures alters the balance of power and may impede the independent development of local governments and societies [1, s.15].

In this regard, the preservation and advancement of the Azerbaijani language play a critical role in counteracting the adverse effects of digital colonialism and safeguarding national culture and economic integrity. As the state language becomes more robustly represented across digital platforms including social media and information retrieval systems it contributes to the protection and continuity of cultural identity, while simultaneously reducing the outflow of linguistically grounded data to foreign systems. The expansion of the state language within digital environments enhances its functional capacity and confers upon it a form of technological agency.

In parallel with the strengthening of the state language, the support and development of local technology companies constitute another essential mechanism for resisting digital colonialism. The creation of digital solutions and software aligned with local linguistic and cultural contexts reduces the market penetration of foreign corporations, thereby fostering domestic economic resilience and limiting external dependency [9].

Furthermore, the development of the state language plays a significant role in enhancing a country’s representation on the international stage. The use of local languages within technological domains facilitates more active participation in the global digital economy and increases resilience against external pressures. This, in turn, enables the formation of independent digital ecosystems.

The Azerbaijani Language in Artificial Intelligence: Challenges, Opportunities and Strategic Implications

Finally, the advancement of the Azerbaijani language in artificial intelligence is not only an economic and cultural imperative but also a matter of national identity. It enhances access to technological knowledge among the local population, enabling citizens to better understand, utilize, and contribute to digital tools and platforms. As a result, dependency on foreign technologies diminishes, and the domestic market becomes increasingly autonomous and sustainable.

Within the paradigm of artificial intelligence, the Azerbaijani language faces not only significant challenges but also substantial opportunities. A considerable portion of these opportunities stems from Azerbaijan's leading role in the contemporary Turkic world, as well as from the position of the Azerbaijani language within the family of Turkic languages.

The Azerbaijani language occupies a central position within the Turkic language family, enabling it to function as a linguistic bridge among related languages. As an intermediary language, Azerbaijani facilitates seamless mutual intelligibility with Turkish, while also serving as a structural conduit to other Turkic languages. From a typological perspective, Azerbaijani may be regarded as a relatively neutral Turkic language that has preserved its agglutinative structure without significant external distortion. Unlike Turkish, which has been subject to substantial European linguistic influence, or certain Central Asian Turkic languages, which have been heavily influenced by Russian, Azerbaijani has retained a comparatively "pure" morphological system. Its root-and-affix structure remains transparent and systematically organized, rendering it particularly suitable as an intermediate model for artificial intelligence and computational linguistic systems [4, s.25].

Furthermore, the Azerbaijani language possesses a rich tradition as a written language. Beyond its function as a medium of everyday communication, it has historically served as the language of classical poetry represented by figures such as Imadaddin Nasimi, Muhammad Fuzuli, and Molla Panah Vagif as well as the language of court administration, diplomacy, journalism, and scholarly discourse across various domains. This extensive literary and intellectual heritage demonstrates the language's capacity to convey abstract concepts, support analytical reasoning, and facilitate the development of specialized discourses in fields such as law, philosophy, and political science. These characteristics are of critical importance for artificial intelligence systems, which rely on semantically rich and structurally complex linguistic data.

In addition, Azerbaijani can effectively function as a cultural and linguistic mediator between Turkish and the Turkic languages of Central Asia. Empirical observations indicate that direct translation from Central Asian Turkic languages into Turkish often results in semantic loss or stylistic rigidity. Azerbaijani, by contrast, shares lexical and structural affinities with both linguistic domains, thereby enabling smoother stylistic transitions and minimizing semantic distortion.

Beyond its translational capacity, the Azerbaijani language also serves as a stabilizing element within the broader Turkic linguistic landscape. It is well recognized that language-related issues in the Turkic world are accompanied by certain sensitivities. The designation of any single Turkic language as dominant or hegemonic may provoke political and cultural resistance among other communities. In this regard, Azerbaijani occupies a uniquely balanced position: it is not demographically dominant, yet it enjoys considerable cultural prestige. Consequently, it does not generate the same degree of political tension and instead functions as a neutral and integrative platform, making it particularly suitable as a common linguistic medium with minimal resistance.

Another strategic advantage of the Azerbaijani language in the context of artificial intelligence lies in its alignment with the economic logic of AI development. Since artificial intelligence systems are largely driven by efficiency and scalability considerations, there is a strong incentive to develop models capable of covering multiple languages with minimal additional cost. Azerbaijani is exceptionally well-positioned in this regard. It demonstrates high compatibility with AI models developed for Turkish and offers a viable pathway for transfer learning into Central Asian Turkic languages. As such, it represents a rational starting point for the development of a unified Turkic artificial intelligence platform. These attributes enable the construction of shared language models that achieve maximum efficiency with limited resources.

Finally, it is essential to emphasize the intrinsic relationship between language, cognition, and algorithmic processing a relationship that is as critical for artificial intelligence systems as it is in the real world. Language is not merely a tool of communication but also a framework for thought. As an agglutinative language with a rich verbal system and complex tense-person relationships, Azerbaijani exhibits strong context-dependent semantic structures. These features introduce both morphological and semantic depth into artificial intelligence models. When effectively leveraged, they can enhance the cognitive flexibility and interpretative capacity of such systems, thereby contributing to more advanced and nuanced forms of machine intelligence.

Properly developed artificial intelligence models, when effectively leveraging the agglutinative properties of the Azerbaijani language, can achieve a more refined understanding of the morphological complexity inherent in natural language. This, in turn, enables the generation of more accurate translations and responses that more closely approximate human cognitive patterns. Moreover, by systematically analyzing a wide range of affixes, AI systems can capture the multifaceted and dynamic nature of language use in Azerbaijani.

The Azerbaijani language is characterized by a high degree of context-dependent semantic richness. A given word or expression may convey multiple meanings not solely based on its direct lexical definition but also in relation to the broader syntactic and textual context in which it appears. Consequently, semantic analysis assumes critical importance in AI-based systems, as meaning construction depends not only on the isolated interpretation of individual lexical units but also on their functional role within context. For instance, the word "gəlib" (derived from the verb "gəlmək," meaning "to come") may express different temporal or aspectual

The Azerbaijani Language in Artificial Intelligence: Challenges, Opportunities and Strategic Implications

nuances depending on contextual usage. The capacity of artificial intelligence to accurately interpret such context-dependent meanings is essential for a more faithful representation of the natural characteristics of the language.

As is well established, Azerbaijani verbs undergo systematic variation according to tense and person, constituting one of the fundamental features of the language. Each verb must be appropriately inflected to reflect both temporal reference and grammatical person. For example, forms such as “gedirəm” (I am going), “gedəcəkdim” (I was going to go), and “gedərəm” (I would go) simultaneously encode tense and person distinctions.

The effective incorporation of these linguistic features into artificial intelligence systems necessitates advanced analytical frameworks and systematic alignment mechanisms. Models must be capable of accurately identifying and interpreting the correspondence between verbal affixes, tense, and person. Proper handling of these relationships enhances the coherence and contextual relevance of AI-generated outputs [7, s.3]. When correctly implemented, these features enable more flexible, precise, and contextually appropriate translations and analyses in Azerbaijani.

Artificial intelligence also plays a crucial role in the digitalization of cultural heritage. As AI systems increasingly operate in Azerbaijani, vast corpora of classical literature, folklore, journalistic texts, and historical documents are transformed into structured digital resources. This process not only facilitates faster, more accurate, and richer analytical capabilities for AI systems but also contributes significantly to the preservation and dissemination of Azerbaijan’s cultural heritage. It ensures the transmission of cultural assets to future generations while simultaneously enabling broader accessibility, including for international audiences.

The reality is that any language absent from technological ecosystems today risks being excluded from the collective memory of the future. At its current stage of development, Azerbaijani represents a language that is open to technological integration and possesses considerable potential in the field of artificial intelligence, yet it remains categorized as a low-resource language. At the regional level, Azerbaijani functions as a strategic and integrative language; at the national level, it serves as an instrument of statehood and identity; and at the technological level, it requires substantial investment and further development [10, s.12].

Although Azerbaijani may not be classified as a mass-market language in the global development of artificial intelligence, it holds strategic importance in terms of cultural sustainability, state sovereignty, and the integration of the Turkic world.

At present, artificial intelligence models are capable of generating relatively coherent and advanced texts in Azerbaijani, which may create the impression that the language is sufficiently represented within technological systems. However, this perception should be approached with caution. From the standpoint of linguistic sovereignty, the current level of representation remains insufficient. AI models generate Azerbaijani-language outputs primarily by drawing on limited open-source textual data, supplemented by statistical and contextual patterns derived from structurally related languages—particularly Turkish and Russian. Consequently, artificial intelligence systems are presently “familiar” with Azerbaijani largely at the level of everyday communication, general analytical writing, and explanatory discourse, which does not constitute full linguistic support.

For the Azerbaijani language to be utilized as an official state language within artificial intelligence systems, a high degree of legal and administrative precision is required. In the absence of such precision, AI systems are not yet capable of providing authoritative interpretations of legislative documents, delivering fully consistent terminological explanations, or making decisions that entail legal responsibility. Within formal legal relations, even a minor misinterpretation of a single term may pose significant risks for the state, underscoring the necessity of achieving a higher level of linguistic accuracy and standardization in AI applications.

Furthermore, contemporary artificial intelligence systems remain largely incapable of adequately accounting for the dialectal diversity of the Azerbaijani language and other sociolinguistic variations. Current models predominantly rely on the standardized literary form of Azerbaijani, particularly texts produced within urban contexts. Consequently, their performance is considerably limited when confronted with regional dialects, colloquial speech, generational linguistic variation, or hybrid registers situated between legal and everyday discourse. However, effective public service delivery requires comprehensive understanding across all such linguistic layers. In addition, the current state of speech technologies in Azerbaijani—including speech recognition, dialect-sensitive language processing, and formal speech synthesis—remains underdeveloped and fragmented, thereby constraining the practical applicability of artificial intelligence systems.

A critical issue lies in the fact that AI-generated Azerbaijani texts are not constructed on the basis of a nationally developed linguistic model but rather through externally developed architectures. The capability of artificial intelligence systems to process Azerbaijani is largely dependent on foreign technological infrastructures, whereas state-level requirements necessitate nationally controlled, legally accountable, and strategically secure systems. In practical terms, although advanced AI models are capable of generating high-quality Azerbaijani texts, the language itself is not embedded within a sovereign technological framework under national control. While such systems may be utilized for routine tasks, they do not provide the level of strategic security required for state functions.

Although artificial intelligence demonstrates a certain level of operational competence in Azerbaijani, it does not treat the language as a primary development priority. As a result, it fails to ensure the degree of precision, accountability, and sustainability required for state-level applications. Critical components such as national data infrastructures, proprietary language models, and independent audit mechanisms remain insufficiently developed.

The Azerbaijani Language in Artificial Intelligence: Challenges, Opportunities and Strategic Implications

It is an unequivocal reality that a language absent from artificial intelligence systems today will likewise be absent from the governance structures of the future [2, s.68]. In the absence of strategic investment in AI development for national and minority languages—including Azerbaijani—the following consequences are likely to emerge progressively:

1. Functional relegation of the language to a secondary status: This does not imply the disappearance of the language; however, in such a scenario, official documentation would increasingly rely on automated translation systems, while analytical processes, decision-support mechanisms, and legal modeling would be conducted in other languages. Consequently, the state language would be reduced to a communicative medium, losing its function as a language of thought.

2. Erosion of legal and terminological systems: Artificial intelligence not only processes language but also generates and standardizes terminology. Without timely investment, legal terminology would be introduced primarily through translation, leading to semantic inconsistencies, fragmentation of legislative language, and the emergence of multiple interpretations for a single concept. Such developments would inevitably result in legal ambiguity and systemic instability.

3. Loss of cognitive grounding in the education system: If AI-driven educational platforms including instructional content, assessment tools, and personalized learning systems operate predominantly in foreign languages, future generations will learn to translate rather than to think in their native language. In this context, the state language would persist as an emotional and бытовой (everyday) medium but would gradually detach from scientific and analytical reasoning, leading to a process of intellectual assimilation.

4. Decline in the functional relevance of the language for future generations: For a language to remain viable, it must generate economic value and be directly linked to career opportunities and technological advancement. If state languages are not sufficiently integrated into AI systems, younger generations may develop uncertainty regarding their practical utility. Although the language may retain formal status, its real-world usage would gradually diminish. Historical evidence suggests that languages do not disappear through prohibition but rather through functional obsolescence.

5. Loss of technological opportunities within the Turkic language family: As previously noted, the Azerbaijani language occupies a structurally central position among Turkic languages, providing a unique opportunity to serve as the foundation for unified AI models across the Turkic world. However, without sufficient investment in its technological development, such models will likely emerge based on other Turkic languages, relegating Azerbaijan to the role of a passive consumer. This would represent the loss of a significant geopolitical and technological opportunity.

6. The imperceptible disappearance of a language: The most critical scenario for any language is not abrupt extinction, but gradual and unnoticed erosion. This process unfolds silently, without explicit prohibition or resistance. At first glance, the language appears to maintain widespread usage and its dominant social position; however, over time—often within a few decades—it becomes evident that, although the language continues to exist in everyday communication, it is absent from technological systems. This absence ultimately leads to its functional obsolescence and eventual marginalization.

7. The emergence of dependency in public administration: If artificial intelligence-based governmental systems—such as tax analytics, social welfare allocation, legal precedent analysis, security frameworks, and risk modeling—are developed in foreign languages, then the underlying decision-making mechanisms will likewise be constructed upon external linguistic and cognitive models. In such a scenario, the state assumes the role of a technological intermediary, which transcends a purely technical issue and becomes a matter of sovereignty.

Given the urgency of this issue, it warrants further elaboration. Artificial intelligence systems operate on the basis of “intelligent” computational frameworks. For a state to maintain full sovereignty, its intelligent systems must be capable of “thinking” and making decisions in the official state language. If these systems function in foreign languages, then the decisions produced are not genuinely those of the state itself but are instead shaped by the cognitive and analytical paradigms embedded within external models.

To illustrate, consider the use of an unfamiliar computational device for performing calculations. If the system’s internal logic is opaque, its programming inaccessible, and its interface presented in an unintelligible language, the user is compelled to input data without fully understanding the process and to accept the output as authoritative. In such circumstances, the computation is performed independently of the user’s reasoning, reducing the user to a mere intermediary in a technologically mediated process.

A comparable dynamic may arise in the application of artificial intelligence within public governance. For example, an AI system used for tax analytics may determine that a particular company presents a high fiscal risk and should be subject to audit. However, if the risk criteria are formulated in a foreign linguistic and economic framework, and if the underlying logic reflects the experience of another country, the resulting decision may not align with the economic realities of Azerbaijan.

Similarly, in the allocation of social assistance, an AI system may analyze household data and conclude that a given family does not meet eligibility criteria. Yet, if the model does not incorporate country-specific factors—such as local family structures, regional living standards, and socio-economic norms—its conclusions may be fundamentally misaligned with reality. In such cases, the state is unable to modify the model and is compelled to implement externally derived decisions, effectively allowing social policy to be shaped by algorithms developed outside national jurisdiction.

The Azerbaijani Language in Artificial Intelligence: Challenges, Opportunities and Strategic Implications

A further example can be observed in legal precedent analysis. An AI system may generate recommendations based on precedent drawn from a different legal tradition, employing translated terminology and reasoning that do not fully correspond to the national legal framework. If judicial authorities begin to align their decisions with such outputs, the legal system risks shifting from an autonomous interpretative framework to one guided by externally constructed models.

These considerations underscore a fundamental principle: language is not merely a system of vocabulary and grammar; it structures categories, establishes priorities, and shapes cognitive frameworks. When systems are constructed in a foreign language, the underlying modes of thought are likewise externalized. Consequently, what is presented to decision-makers is not an internally generated conclusion but a translated outcome derived from an alien cognitive structure.

Admittedly, in cases involving universally defined and logically deterministic operations—such as basic arithmetic (e.g., $2 + 2$)—the language in which an AI system operates is largely irrelevant. However, public institutions rarely function on the basis of such unambiguous computations. Questions such as “Is this individual high-risk?”, “Is this household economically disadvantaged?”, “Is this company suspicious?”, or “Is this behavior dangerous?” do not yield definitive answers, as they involve evaluative judgment rather than purely quantitative calculation.

For instance, consider a social welfare assessment in which a household has a monthly income of 600 AZN. The central question—“Is this household genuinely impoverished?”—cannot be answered in isolation. If the household resides in Baku, the answer may be affirmative; if it resides in a rural region, the answer may differ; if the household includes five children, the conclusion may again shift; whereas a single-child household would yield yet another interpretation. Although the numerical value remains constant, the decision is contingent upon contextual factors, which are themselves shaped by culturally embedded linguistic cognition.

If the “thinking” framework of an artificial intelligence system is detached from the local linguistic and socio-cultural environment, and if its evaluative criteria are derived from external contexts, then even ostensibly correct answers may, in practice, be inappropriate or erroneous.

Artificial intelligence systems do not, strictly speaking, make decisions based directly on language; however, prior to decision-making, they transform reality through a sequential process—life → words → categories → decision logic—thereby converting all inputs into linguistic representations. In this process, the world is not initially partitioned through numerical values but through conceptual and lexical structures.

When artificial intelligence relies on foreign linguistic infrastructures, it constructs such categories using external conceptual frameworks—for instance, “low income,” “large household,” or “unstable employment.” These terms are subsequently translated into Azerbaijani, yet their semantic parameters remain anchored in the socio-economic realities of other countries. Although the lexical forms may appear equivalent, their underlying meanings are not calibrated to the Azerbaijani context. For example, in one society, a two-child household may be considered “standard,” whereas in another it may be classified as “small.”

In summary, if sufficient investment is not directed toward the technological development of the Azerbaijani language—and other national languages—within artificial intelligence systems, these languages risk gradual exclusion from future governance frameworks. Such a trajectory may result in technological dependency, the reduction of language to a cultural artifact rather than a functional medium, and the erosion of its role in decision-making processes. In this context, investment in language is, fundamentally, an investment in the future sovereignty of the state.

CONCLUSION.

The role of the Azerbaijani language in the era of artificial intelligence extends far beyond a purely linguistic or technological dimension, encompassing critical issues of sovereignty, cultural continuity, and strategic autonomy. While the language currently faces significant structural and technological limitations—particularly as a low-resource language—its unique position within the Turkic language family, rich literary heritage, and morphological characteristics offer substantial opportunities for integration into advanced AI systems. However, without systematic investment in national language resources, including corpora, annotated datasets, domain-specific lexicons, and locally trained models, Azerbaijani risks gradual marginalization within digital ecosystems and exclusion from future decision-making infrastructures. The development of AI systems capable of operating in the official state language is therefore not merely a matter of technological advancement, but a prerequisite for maintaining independent governance, ensuring accurate and context-sensitive decision-making, and preserving national identity in an increasingly algorithm-driven world. Consequently, investment in language technologies should be regarded as a strategic priority, directly linked to the long-term sustainability and sovereignty of the state.

REFERENCES

- 1) Avila Pinto, R. (2018). Digital sovereignty or digital colonialism, *SUR-Int'l J. on Hum Rts.*, p. 15
- 2) Bird, S., Klein, E., Loper, E. (2009). *Natural Language Processing with Python*. Sebastopol: O'Reilly Media. 504 p.
- 3) Bori, A.M. (2024) Data colonialism: a contemporary manifestation of old practices or an oldfashioned term for a contemporary phenomenon?, *The bulletin of the Graduate School, Soka University*, (45), 211-244
- 4) D.Jurafsky, James H. Martin. *Speech and Language Processing*. <https://web.stanford.edu/~jurafsky/slp3/>

The Azerbaijani Language in Artificial Intelligence: Challenges, Opportunities and Strategic Implications

- 5) Eisenstein, Jacob (2019). Introduction to Natural Language Processing. The MIT Press. p. 1
- 6) G.Khensous, K.Labed, Z.Labed (2023). Exploring the evolution and applications of natural language processing in education. Romanian Journal of Information Technology and Automatic Control, Vol. 33, No. 2, 61-74
- 7) J.Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. (2018) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, <https://www.researchgate.net/publication/328230984>
- 8) Mahmudov, M. (2024). Linguistic problems of artificial intelligence. Baku: "Elm və Təhsil"https://ecspm.org/wp-content/uploads/2018/12/EC-Comprehensive-Approach-to-LL_2018.pdf
- 9) Manning, C. D., Schütze, H. (1999). Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press. 680 p.
- 10) R.Alguliyev, Y.Jafarov, R.Aliguliyev (2024). ICT problems of the Azerbaijani language, the Azerbaijani language problems of ICT. Problems of Information Society, 2024, vol.15, no.2, 3-13