
www.jpis.az

Linguistic sovereignty in virtual space: language ecosystem, problems and solutions

Afruz Gurbanova¹, Mehriban Baghirova²

Institute of Information Technology, B. Vahabzadeh str., 9A, AZ1141, Baku, Azerbaijan^{1,2}

¹afruz1961@gmail.com, ²mehriban.amea@gmail.com

[1orcid.org/0000-0002-2274-8015](https://orcid.org/0000-0002-2274-8015), [2orcid.org/0009-0009-5108-5001](https://orcid.org/0009-0009-5108-5001)

ARTICLE INFO

<https://doi.org/10.25045/jpis.v17.i1.11>

Article history:

Received 12 September 2025

Received in revised form

15 November 2025

Accepted 25 January 2026

Keywords:

Linguistic sovereignty
Digital ecosystem
Artificial intelligence
National language technologies
Content production
Digital language policy
Technological independence

ABSTRACT

Against the backdrop of globalization and digital transformation in the modern era, linguistic sovereignty has become a strategic cybersecurity issue, extending beyond the traditional paradigm of state security. As languages migrate from physical space to virtual ecosystems, national languages face the hegemonic influence of global tech giants, and, especially in the case of dominant languages, artificial intelligence systems. The lack of high-quality text corpora and natural language processing technologies for national languages in virtual environments leads to linguistic asymmetries in artificial intelligence systems, thereby limiting the functionality of national languages on digital platforms. Protecting linguistic sovereignty in the digital environment is crucial to ensuring the functional existence and sociocultural sustainability of national languages. This article analyzes the functional independence and sustainability of languages in the context of the digital ecosystem, identifying the key aspects of linguistic sovereignty. Conceptual approaches to ensuring and developing linguistic sovereignty in the digital environment are analyzed, and a multi-level model of linguistic sovereignty in this context is proposed. The main threats to linguistic sovereignty in the virtual space are identified, and the need to develop an ecosystem of national languages to address them is emphasized. The possibilities and key components of developing a language ecosystem on a digital government platform are identified. Recommendations are provided for expanding the functionality of national languages in the digital space and ensuring digital sovereignty. The development of a sustainable ecosystem of the Azerbaijani language on a digital government platform is proposed as a promising direction.

1. Introduction

Globalization and digital innovation are fundamentally altering the relationship between language, technology, and society. These changes directly impact the functional position, security, and resilience of language in the information environment. Linguistic sovereignty refers the right of a state to independently determine the status of its official language and regulate its use in society, as well as to protect and develop the languages of all peoples living in the country. Language is not only a means of communication but also an institutional bearer of national identity,

cultural heritage, and collective historical memory. Therefore, the protection of linguistic sovereignty is regarded as a strategic mechanism for ensuring a state's cultural independence and internal security.

The classical Westphalian model of sovereignty – the concept of international sovereignty that emerged after the Peace of Westphalia in 1648 and is based on the principles of territorial integrity and non-interference in the internal affairs of states – has faced a number of limitations in the modern era (Hayman et al., 2006; Paris, 2020). In the post-sovereign stage, a state's de facto independence is measured by its resistance to

external hegemonic influence, manifested on linguistic platforms.

In the modern era, linguistic sovereignty faces new challenges from digital technologies and global artificial intelligence (AI) platforms. For this reason, states need to reconsider their strategies for protecting and regulating linguistic sovereignty in cyberspace. The approaches to cyber sovereignty presented in this context form the theoretical basis for linguistic sovereignty in the digital space (Hung, 2025; Barrinha et al., 2022). According to the researchers' analysis, cyber sovereignty is not limited to a defensive function but also includes control mechanisms at the levels of physical infrastructure (networks, servers, etc.), code (natural language processing (NLP) technologies for the national language), and information (high-quality language corpora, resources, etc.). This approach demonstrates that digital language policy, in addition to being a cultural choice of the state, is becoming an integral part of national security strategy.

Italian philosopher Luciano Floridi interprets digital sovereignty as extended control exercised not only over physical infrastructure but also over software, standards, processes, and services. His conceptual approach emphasizes the importance of the institutionalization of language not only at the information level but also within digital platforms (e.g., national-language software, AI services, etc.) to ensure linguistic sovereignty (Floridi, 2020). The global scale of digital transformation and the emergence of large language models (LLMs) (GPT, Claude, Gemini, etc.) based on AI technologies pose new threats to small and medium-sized languages. Thus, studying the theoretical and practical foundations of ensuring linguistic sovereignty in the cyber environment has become a priority area of contemporary academic discourse.

2. Related work

The preservation, development, and integration of national languages into the digital environment have become a priority both in academic circles and in the context of national policy. This study analyzes research covering international and regional experiences, theoretical foundations, and technological solutions in this field.

The study (Alguliev et al., 2025) examines the protection and integration of the Azerbaijani language within the e-government platform in a digital and globalized environment. The paper

demonstrates that the creation of language corpora, the implementation of various language services, and the use of linguistic technologies are necessary for the protection of the national language and its development in a unified information space. The research presents a conceptual model of the Azerbaijani language ecosystem on the electronic government platform and proposes strategies for protecting linguistic sovereignty.

(Kahiraba, 2025) examines the impact of globalization on linguistic and cultural sovereignty in the African context, specifically in Mali. Drawing on empirical data, the author demonstrates that the hegemony of global languages weakens the social functions of local languages and narrows the scope for expressing national cultural identity. The study concludes that linguistic sovereignty can be protected through a combination of national policy, education, and public awareness.

(Kraus, 2018) offers a scholarly analysis of the interaction between languages and the challenges of national language independence in the era of globalization. The author argues for the need to actively promote models of linguistic sovereignty that go beyond the traditional concept of controlling the cultural and linguistic rights of majorities and minorities.

(Joshi et al., 2020) examines the inequality in resource distribution between languages in the field of artificial intelligence and digital translation technologies (translation systems, speech recognition, etc.). Because digital translation systems are primarily built on dominant languages, underresourced languages are excluded from the digital environment, which seriously impacts linguistic diversity. According to the authors, equitable resource distribution, data collection initiatives, and the expansion of multilingual models are essential for increasing linguistic inclusivity.

The paper (Tan et al., 2024) examines issues of language stability and change in the context of linguistic sovereignty in the digital environment and shows the opportunities and risks that digital technologies create for the preservation of minority languages.

(Siminyu et al., 2021) The AI4D (Artificial Intelligence for Development) model is considered a successful concept that can be applied to digital language policy at the national and regional levels. This program is based on the creation of open resources for African languages and the application

of AI technologies (speech recognition, machine translation, etc.) in these languages.

(Scannell, 2007) presents a methodology for creating automated corpora for resource-limited languages using internet resources. This approach is considered an important step in documenting digital languages and generating source data for language technologies.

(Anastasopoulos et al., 2020) focuses on ethical and equitable principles of representation in the development of multilingual translation models. The authors emphasize that technological development must ensure equal opportunities for all languages, not just dominant ones. This conceptual approach is directly related to the digital equality aspect of linguistic sovereignty.

The analyzed sources emphasize that the state of national languages, the narrowing of their functional scope, and the maintenance of linguistic sovereignty constitute a foundation for national identity and cultural independence in the era of globalization and digitalization. In this regard, the creation of language corpora and data collection initiatives are essential for the application of linguistic technologies. The use of machine translation, speech recognition, and other language services in the national language is crucial. Protecting linguistic sovereignty requires concerted action by national policy, education, and public awareness. Based on the results of the conducted research, the integration of the Azerbaijani language into the electronic government platform, the creation of language corpora, the application of linguistic technologies, and the development of concepts and strategies in this area can be considered essential.

It should be noted that the sources reviewed focus primarily on resource shortages and the use of technology, while the issue of dependence on external technological platforms in building a national language ecosystem has not been analyzed in depth. Language technologies are presented as a socio-academic project. The relationship between national security, personal data protection, and linguistic sovereignty in the creation of national language corpora is insufficiently addressed in the studies.

Thus, against the backdrop of globalization and digital transformation in the modern era, linguistic sovereignty has become a strategic cybersecurity issue that extends beyond traditional nation-state security paradigms. As the plane of language existence shifts from physical space to a virtual ecosystem, national languages face the hegemonic influence of global tech giants and, especially, artificial intelligence systems in dominant languages.

In this environment, the lack of high-quality text corpora and natural language processing technologies for national languages leads to linguistic asymmetry in AI systems, limiting the functionality of the national language on digital platforms. At the same time, the abundance of content in dominant languages narrows the scope of national language use, directing users to external resources, posing a serious threat to cultural diversity and national identity. Consequently, the classical concept of sovereignty does not fully encompass the challenges of the digital age. A state's true independence is currently measured by its control over its language, its digital infrastructure, software, and artificial intelligence services.

Ensuring the independent, functional existence of the national language in the virtual ecosystem has become a necessary and strategic priority in the context of national security, cultural sustainability, and the protection of information equality. The main objective of this study is to analyze current challenges related to ensuring linguistic sovereignty in the virtual environment, identify problems and solutions based on scientific and theoretical approaches, and propose an interdisciplinary methodology for the digital development of national language resources.

To achieve this goal, this study examines the concept of linguistic sovereignty, its essence, and identifies its key aspects. It systematically examines threats to state linguistic sovereignty in the digital environment and proposes a multi-layered model of linguistic sovereignty in the context of cyber sovereignty. The possibilities and key components of developing a linguistic ecosystem on a digital state platform are identified. Recommendations are provided for expanding the functionality of national languages in the digital space and ensuring their digital sovereignty. The establishment of a sustainable Azerbaijani-language ecosystem on a digital state platform is proposed as a promising strategic direction.

3. Materials and methods

In addition to explaining the essence of linguistic sovereignty, the study also analyzes the impact of digital platforms, algorithmic governance mechanisms, artificial intelligence-based systems, and the global communication environment on language. The theoretical literature and contemporary conceptual approaches used in the study are analyzed. Given the highly dynamic and fluid nature of the digital environment, the study

offers a comparative analysis of information obtained from various sources, as well as legal and regulatory frameworks.

The research methodology combines theoretical and conceptual analysis, as well as an analytical assessment of real-world processes in the digital environment. This approach allows for a more comprehensive understanding of the challenges and opportunities facing linguistic sovereignty in the era of globalization and digital transformation, and serves to scientifically substantiate strategic decisions regarding linguistic security.

3.1. The concept and essence of linguistic sovereignty

The concept of linguistic sovereignty has emerged as a conceptual framework aimed at protecting the functional independence of national languages in the context of contemporary globalization and digitalization. This concept expresses the control of the state and society over the national language, its ability to ensure its independent existence and development in the digital, technological, and communicative spheres. The globalization of the digital ecosystem highlights the functions of language not only as a means of communication but also as a mechanism for knowledge production, national identity, and social legitimacy.

Language is a fundamental component of collective identity, cultural heritage, political independence, and state sovereignty. For this reason, the concept of linguistic sovereignty has been widely examined across several academic fields, such as linguistics, sociology, language policy, and international law.

The concept of linguistic sovereignty found its theoretical basis in a broader context as an institutional defense mechanism of states against the pressures of linguistic hegemony created by the theory of linguistic imperialism put forward by Robert Phillipson (Phillipson, 2008).

As the fundamental force of state sovereignty, this concept ensures linguistic space on two main levels (Eckes, 2015):

- At the level of internal sovereignty, linguistic sovereignty is ensured by the constitutional and linguistic legislation of the state. This includes defining the normative framework for the use of the state language in key functional areas such as public administration, education, media and the judiciary.

- From the point of view of external autonomy, linguistic sovereignty allows the state to implement mechanisms of protection against linguistic hegemony created by globalization and transnational media, and serves as a defense against external cultural influence.

Digital language policy is not only a state's cultural choice but also an integral part of its national security strategy. The functional weakening of a language in the digital environment should be considered as a critical threat, potentially leading to the decline of national identity and, consequently, the undermining of the state's internal sovereignty.

Linguistic sovereignty covers a number of key areas that ensure the status, development, use and protection of a language (Gazzola, 2016).

- Ensuring the active participation of language in international diplomacy, organizations, and digital ecosystems—the internet, social platforms, and artificial intelligence environments;
- Protecting the right to control grammatical, spelling, and lexical norms of language without any external interference;
- Unrestricted use of language in administrative, educational, media, judicial, and economic spheres;
- Implementing appropriate policies at the state level to protect dominant languages from assimilation pressure;
- Active use of language in educational and scientific systems, the creation of new terminology, and its transmission from generation to generation;
- Constitutional recognition of language, ensuring its legal status and linguistic rights;
- Widespread use in the media and public communication;
- Collection, preservation, and effective use of language resources in digital format in artificial intelligence and digital technologies;
- Promoting educational activities that enhance the cultural and social value of language in society.

These aspects complement each other and create the necessary conditions for the survival of a language and the preservation of a community's cultural identity. The acceleration of digital transformation is changing the scope of traditional aspects of linguistic sovereignty. The existence of language in the digital environment has become not only a socio-cultural issue but also a key component of technological and information

security. This necessitates a rethinking of the concept of linguistic sovereignty in the digital space and a systematic analysis of its relationship with cyber sovereignty, data sovereignty, and artificial intelligence.

Linguistic sovereignty is based on the generally accepted principles of state sovereignty in international law. A number of international documents support the right of states to independently determine their language policies. The UNESCO Universal Declaration on Cultural Diversity (UNESCO, 2001) calls on states to protect and promote their languages, recognizing language as an integral part of national cultural heritage. An approach based on international law strengthens linguistic sovereignty as a legitimate and strategic instrument for protecting against linguistic hegemony in the context of globalization.

3.2. Problem solution

Since the violation of linguistic sovereignty in the virtual space is caused by such fundamental problems as the narrowing of the language's functional domain, technological dependence, and a shortage of national content, addressing this issue requires the coordinated application of multi-level and integrative strategies. To achieve this, technological independence must first be strengthened. Thus, to counteract global linguistic hegemony, high-quality language corpora should be created in the national language, large language models should be adapted to the national language, and local control over natural language processing technologies should be ensured. The next step should be the creation of high-quality digital content (text, audio, video) in the native language. This will help eliminate the content shortage. Technological and content-related efforts must be supported by digital language policy and institutional regulation. This includes the creation of a regulatory framework that ensures the sovereignty of information (data) in the national language, regulates the use of language in the digital environment, and promotes algorithmic equality. These comprehensive solutions ensure the sustainability of the national language in the digital ecosystem and strengthen technological independence.

3.3. Digital linguistic sovereignty

In the context of the rapid digitalization of economic, social, and cultural life in the modern era,

preserving linguistic diversity and ensuring linguistic sovereignty are of paramount importance. The widespread use of global digital platforms, the development of artificial intelligence systems, and the increase in cross-border information flows create both significant opportunities and serious challenges for language. These challenges are particularly acute for languages with limited resources and limited geographic distribution (Kelly-Holmes, 2019).

The concept of digital linguistic sovereignty refers to the ability of a people or state to freely use and control their language through digital technologies. This encompasses the right to unrestricted language use in the digital world, as well as the right to communicate, teach, learn, and participate in the digital economy in one's native language.

In the era of digital transformation, language self-determination is linked to the creation of technical and institutional infrastructure for their survival and development in the digital space. This includes the development of keyboard systems in the native language, software localization, and ensuring the language's presence online. In this area, it is important to utilize not only technical resources but also cultural, economic, and social tools. People's access to digital resources in their native language enhances their digital inclusion, literacy, and equal opportunities for participation.

The main indicators for ensuring linguistic sovereignty in the digital environment are the following: (Kornai, 2013; Moshagen et al., 2024; Phillipson, 2008):

- **Online language presence:** ensuring the functional presence of the language in search engines, social media and digital platforms.
- **Digital language tools:** availability of interfaces, keyboard layouts, dictionaries, translation tools, and AI-based software in the national language.
- **Education and digital literacy:** Promoting digital education in the mother tongue and increasing the ability to use information technology in that language.
- **Protection from the influence of dominant languages:** ensuring the protection of local languages from assimilation and technological marginalization.
- **Language use in public services:** application and standardization of native language in digital public services.

Digital language policy and technological sovereignty play an important role in the following areas in terms of national security and identity in the modern era:

- Digitization of audio, video, and written sources for language preservation.
- Expanding the functional scope of languages through e-learning, social media, and digital platforms.
- Creation of digital content in the native language in music, literature, and fine arts.
- Expanding access to local and international information in the native language.
- Ensuring language security through national technological solutions.
- Protecting the integrity of the national information space through language.
- The impact of using the national language in the digital world on social cohesion and national identity.
- Developing technological products in the national language.
- Adapting artificial intelligence tools to languages with limited resources.
- Developing new forms of expression and stylistic models.
- Promoting communication and mutual understanding between speakers of different languages.
- Increasing youth interest in their native language through social media.
- Introducing digital products and services in the national language into economic circulation.
- Representing national languages in the global digital space.
- Ensuring digital equality by translating information into native languages.

These aspects strengthen the resilience and social position of language in the digital environment and play a vital role in ensuring national identity and technological independence. It is essential to develop a digital language policy and ensure the full participation of national languages in the digital ecosystem. Fig. 1 presents a multi-layered model of digital linguistic sovereignty.

Digital autonomy of the individual. Digital autonomy of the individual constitutes the entry level of the model. It encompasses the individual's right to freedom of expression in the digital space and to use online communications, education, and services in their native language. The primary goal is to prevent linguistic discrimination and ensure equal participation in social and digital discourses in their native language. Relevant UNESCO declarations (UNESCO, 2001; UNESCO, 2023) and initiatives such as the Council of Europe's Universal Framework of Competences (Council of

Europe, 2020) support the implementation of linguistic rights in the digital environment within the framework of cultural rights.

Digital national content production (creation and distribution of digital content). Content sovereignty ensures the vitality and sustainability of a language in the digital environment. Digital language vitality refers to the use of the language in the digital environment (web, social media, mobile apps, etc.), attracting new users, and expanding its functionality. The main goal is to expand the production of high-quality and accessible digital content in the native language in various formats (text, audio, video). Indexing content in search engines increases the social influence of the language and creates the basis for the development of platforms that facilitate online knowledge exchange (Rehm et al., 2020).

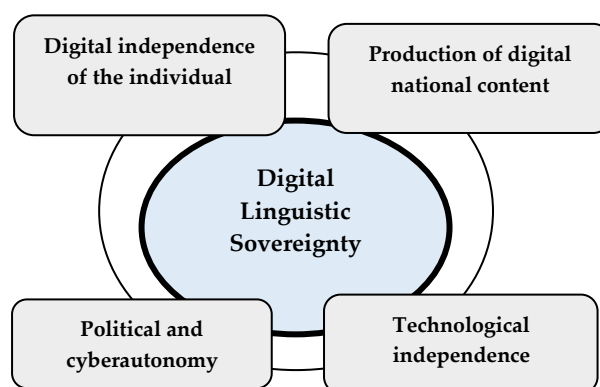


Fig. 1. Multilevel model of linguistic sovereignty in the digital environment

Technological independence (language support through digital technological infrastructure). To ensure functional language support through digital technological infrastructure, it is necessary to develop support for national languages in software, such as keyboards, dictionaries, and grammar tools. Most importantly, adapting artificial intelligence and digital technological independence technologies to language is a key factor in ensuring digital equality. Supporting native languages in recommender systems increases content accessibility (Bella et al., 2023).

Political and Cyberautonomy (Institutional Framework). This level forms the institutional and normative foundations for the governance of the digital language space. This includes the development of digital language policies, the implementation of language standards on government portals, and the development of language-related legislation. Initiatives by international organizations such as ICANN and

UNESCO in the area of the International Domain Name System (IDN) contribute to strengthening linguistic sovereignty globally (Bradshaw et al., 2018). Furthermore, data sovereignty, which ensures full control over digital information created in one's native language, is critical in this context.

3.4. The main threats to linguistic sovereignty in the digital environment

Having analyzed the essence of linguistic sovereignty and its institutional connections to international law, and state sovereignty, it is necessary to identify the main threats and strategic challenges it faces in the modern era. These threats, which threaten the functional existence and sociocultural sustainability of national languages in the digital environment, can be broadly summarized into three analytical categories:

- Pressure from global linguistic hegemony;
- Linguistic asymmetry in artificial intelligence systems;
- Lack of digital national content.

These categories define the status of language in the digital ecosystem (fig. 2).

The pressure of global linguistic hegemony. This threat stems from the superiority of dominant world languages (especially English) in economic, technological, and scientific spheres. Global linguistic hegemony is manifested in two principal aspects:

- **Narrowing of the functional sphere of the national language:** The dominance of dominant languages in the spheres of international communication, higher education, scientific research and high technology leads to the gradual displacement of the national language from these functional areas. As a result, the use of the national language is mainly limited to everyday and traditional cultural spheres, which weakens its social and functional significance.
- **Decrease in lexical and terminological richness:** The lack of systematic development of a national terminological base limit the ability of the language to meet modern scientific and technological requirements. This weakens the ability of the national language to function in highly functional areas (e.g., in engineering and science) and reduces its overall status.

Language asymmetry in AI systems. Language asymmetry in AI systems is one of the newest and most serious threats to linguistic sovereignty. Because AI systems are data-driven, small and medium-sized languages face data shortages (Chen et al., 2024). This asymmetry poses three main risks:

- **Data scarcity:** The lack of large, labeled, and diverse datasets needed to train high-quality AI systems in a national language makes it difficult to build models for local languages;
- **Algorithmic discrimination:** Because global AI models are built primarily on data in dominant languages, responses and translations to queries in national languages often leave much to be desired. This creates inequality in access to digital services for speakers of national languages;
- **Risk of digital destruction:** Overreliance on AI-based translation and content creation tools increases the risk of violating the lexical and grammatical norms of the national language. This could lead to "standardization" and distortion of the language structure under the influence of AI.
- **Lack of digital national content.** Language sovereignty in the digital space depends on its representation through high-quality, diverse, and engaging content. The limited availability of digital content in the national language forces users, especially young people, to satisfy their information needs through sources in dominant foreign languages. This trend leads to the marginalization of the national language in the digital ecosystem. These threats interact and threaten linguistic sovereignty throughout the entire digital value chain.

In the modern era, defending linguistic sovereignty requires understanding the dual impact of globalization. On the one hand, globalization accelerates language shift and the weakening of minority languages through the uneven advancement of dominant languages. On the other hand, global technological integration creates significant opportunities for language revitalization in the digital environment. This facilitates the survival of languages through digital archiving, online learning, and communication (Gurbanova, 2022). Therefore, language warfare should be waged not through universal models, but through context-specific strategies — that is, strategies that align with the social and technological realities of each linguistic community (Bird, 2020; Kornai, 2013).

Ensuring linguistic sovereignty must be carried out in three interrelated directions:

- **Technological integration and language viability in the digital environment:** ensuring the functional viability of language in the digital environment. This includes not only the creation of translation tools but also the creation of national language corpora, the development of digital translation systems, and the elimination of the problem of "data scarcity."

- **Sociocultural sustainability:** measures that promote the transmission of language from generation to generation and encourage the creation of digital content in the native language in various fields play a key role.
- **Institutional and regulatory framework:** development of digital language policy, creation of terminology banks and ensuring standardized use of the national language in digital products.

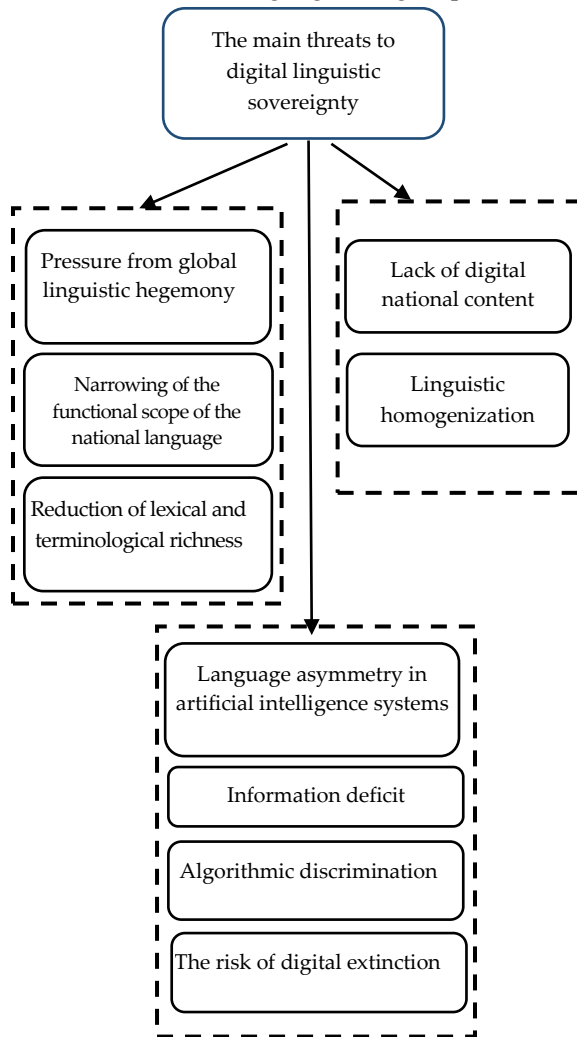


Fig. 2. Diagram of the main threats to linguistic sovereignty in the digital environment

The synergy of these three directions ensures not only the preservation of the national language but also its development as a competitive language in the digital age.

Developing national language technologies. A fundamental pillar of digital sovereignty. The primary mechanism for ensuring linguistic sovereignty in the digital age is the equal representation of the national language in the technological ecosystem. The development of natural language processing systems is a key tool in this direction. The development of technologies

such as speech recognition and synthesis, text analysis, and automatic translation into the national language will transform language from a passive means of communication into an active technological interface (Aliguliev et al., 2018). This approach aligns with the tenets of the theory of technological determinism: the social status of a language is directly related to its representation on technological platforms. To achieve this, corpus planning is a priority. The creation of high-quality, large-scale, and tagged text and speech corpora is key to overcoming the problem of "data deficit" in the national language and preventing algorithmic bias. In this regard, it is necessary to encourage investment in open-source national AI models through public-private collaboration. This will enable local communities to participate in knowledge production.

Stimulating the production of digital content in the native language. The digital sustainability of a national language is closely linked to its functional viability and transmission from generation to generation. A language's survival in the digital space depends not only on technological support but also on user demand. Therefore, it is necessary to encourage the creation of high-quality and diverse content (videos, blogs, podcasts, scientific resources, etc.) in the native language (Galla, 2016). Functional diversity of content promotes the active use of the language not only in everyday life and entertainment, but also in fields such as science, technology, and economics. It is important to support content producers through programs and incentive mechanisms at the public and private levels. These measures strengthen the national media ecosystem.

Developing a Digital Language Policy: Normative and Institutional Support. Effectively ensuring linguistic sovereignty requires an institutional approach. Since existing language policies do not take into account digital challenges and the impact of AI, it is necessary to integrate digital components into these documents. The key areas are as follows:

- **Terminological standardization:** ensuring the creation and use of a unified terminological base (terminology banks) in the national language (Aliguliev et al., 2018).
- **Localization of digital technologies:** ensuring the full and high-quality implementation of the national language in all digital products and services commissioned by the government as a regulatory requirement.

- **Language use in online learning:** developing mechanisms to ensure the dominance of the native language in digital resources at all levels of education.

At the same time, it is essential to implement educational programs that promote the ethical and normative use of language on digital platforms. These measures confirm that linguistic sovereignty is an integral element of national security and governance.

To protect linguistic sovereignty in the era of digital transformation and globalization, it is essential to develop all components of the national language ecosystem in a unified and coordinated manner. This includes the provision of software in the national language, artificial intelligence and natural language processing services, language support on digital platforms, and the functional use of the language in the educational and media ecosystem. This approach guarantees equal participation in public and digital discussions in one's native language, prevents linguistic discrimination, and protects national languages in the digital space.

3.5. The main components of the national language ecosystem

The development of a language ecosystem is one of the key tasks in preserving, developing, and adapting a language to contemporary challenges. A language ecosystem encompasses all factors, environments, and resources that influence its functioning and development—the education system, media, digital technologies, and everyday communication (Aliguliev et al., 2021).

There are several problems that hinder the formation of a language ecosystem:

- *Dominance of dominant languages (e.g., Russian and English) in certain fields:* In academia, IT, and business, the significant influence of dominant languages is sometimes observed. This leads to the creation and use of content, documentation, and communication in a language other than the native language. This limits the use of the native language in such important fields.
- *An underdeveloped digital environment:* digital content in the native language, including terminology databases, software, online courses, and high-quality web resources, is limited. This creates barriers to language use in the context of increasing digitalization.

- *Problems with terminology and standardization:* the lack of a unified and generally accepted terminology in the native language in some areas creates challenges in scientific research, professional communication and education.
- *Low share of high-quality content in the national language in video games and other popular media formats.*
- *Insufficient attention to linguistic culture in everyday life:* the influence of borrowed words, the use of slang, and non-compliance with literary language norms in everyday communication can weaken the position of the native language.
- *Quality of Teaching:* In some cases, the quality of teaching of the native language, especially as a second language, may be insufficient, which affects the level of language proficiency among the population.

To address these issues and implement initiatives related to the preservation and use of language on the digital government platform, the implementation of relevant programs and projects should be carried out in stages (Gurbanova, 2015; Hornberger et al., 2018; Haddow et al., 2022):

Enrichment of digital infrastructure and content base:

- *Creation of a comprehensive terminology database and online dictionaries:* Development and maintenance of an open, accessible and continuously updated terminology database for all areas of science and technology;
- *Support for the development of software and digital products in native languages:* stimulating the localization of existing programs and the creation of new ones oriented towards national languages;
- *Development of the national language space on the Internet:* creation of high-quality websites, blogs, educational platforms and multimedia content in the national language;
- *Use of artificial intelligence technologies:* The use of artificial intelligence technologies to develop machine translation systems, speech recognition and synthesis in national languages in order to improve the accessibility and usability of language in the digital environment.

Strengthening the role of language in education and science:

- *Priority of the national language in education at all levels:* support and development of educational programs in the national language.

- *Stimulating scientific research in the national language*: creating conditions for publishing scientific articles in the state language, holding conferences and exchanging experiences.
- *Development and application of modern methods of teaching the national language*: improving the qualifications of teachers and creating attractive teaching materials.

Strengthening the role of the national language in the media and culture:

- *Supporting the production of high-quality media content*: encouraging the creation of films, television programmes, music and video games in the national language.
- *Expanding the presence of the national language on international platforms*: localizing popular services and promoting content created in the national language globally.
- *Promoting the use of the national language in everyday life*: conducting information campaigns aimed at improving linguistic culture and the conscious use of the literary language.

State support and legislative framework:

- *Effective implementation of legislation on the state language*: monitoring compliance with the norms of use of the national language in official documentation, advertising, and announcements.
- *Development and implementation of national programs*: preparation of long-term strategies and plans for the development of the national language.
- *International cooperation*: exchange of experience with other countries on issues of language policy and development of language ecosystems.

Thus, the development of a fully-fledged language ecosystem on a digital state platform is a long and multifaceted process that requires the joint efforts of the government, academia, educational institutions, businesses, and every native speaker. Only a comprehensive approach will ensure sustainable language development in a changing world. Fig. 3 illustrates the main components of a national language ecosystem:

The Unified Language Platform, which constitutes the core of the national language ecosystem, is the hub for storing and processing

all language resources. It includes a unified terminology database, language models, and validation tools that ensure accuracy and consistency. The Language Services functional unit is one of the key components, directly interacting with users and other systems. All these services utilize the resources of the language platform. The Integration functional unit facilitates interaction between the national language ecosystem and other digital platforms, such as the Digital State. The Analytics and Management functional unit manages, monitors, and optimizes all processes within the national language ecosystem. This ensures the collection and processing of information on system operation, the evaluation of the effectiveness of language models and services, and resource management.

The national language ecosystem emerging in the digital environment serves as a complex infrastructure of strategic importance for ensuring linguistic sovereignty.

This model, in which a unified language platform plays a central role, enables the collection, management, and use of national language resources in a unified information environment, a crucial function for maintaining control over the language and reducing external technological dependence. Technological components such as terminology databases, text corpora, knowledge bases, automatic translation, speech recognition, and speech synthesis strengthen the digital stability of the national language and ensure its full functionality in global information systems.

Online translators, voice assistants, content generation tools, and language audit mechanisms, integrated into a suite of services and applications, ensure the regulatory uniformity of the language, standardization, and quality control of its use in the digital sphere. Integration with the digital state through automated document management, API interfaces, and unified standards further strengthens the legal, functional, and technological sovereignty of the language at the level of public administration. The synchronous operation of these components allows the national language to maintain an independent development trajectory in the era of digital transformation

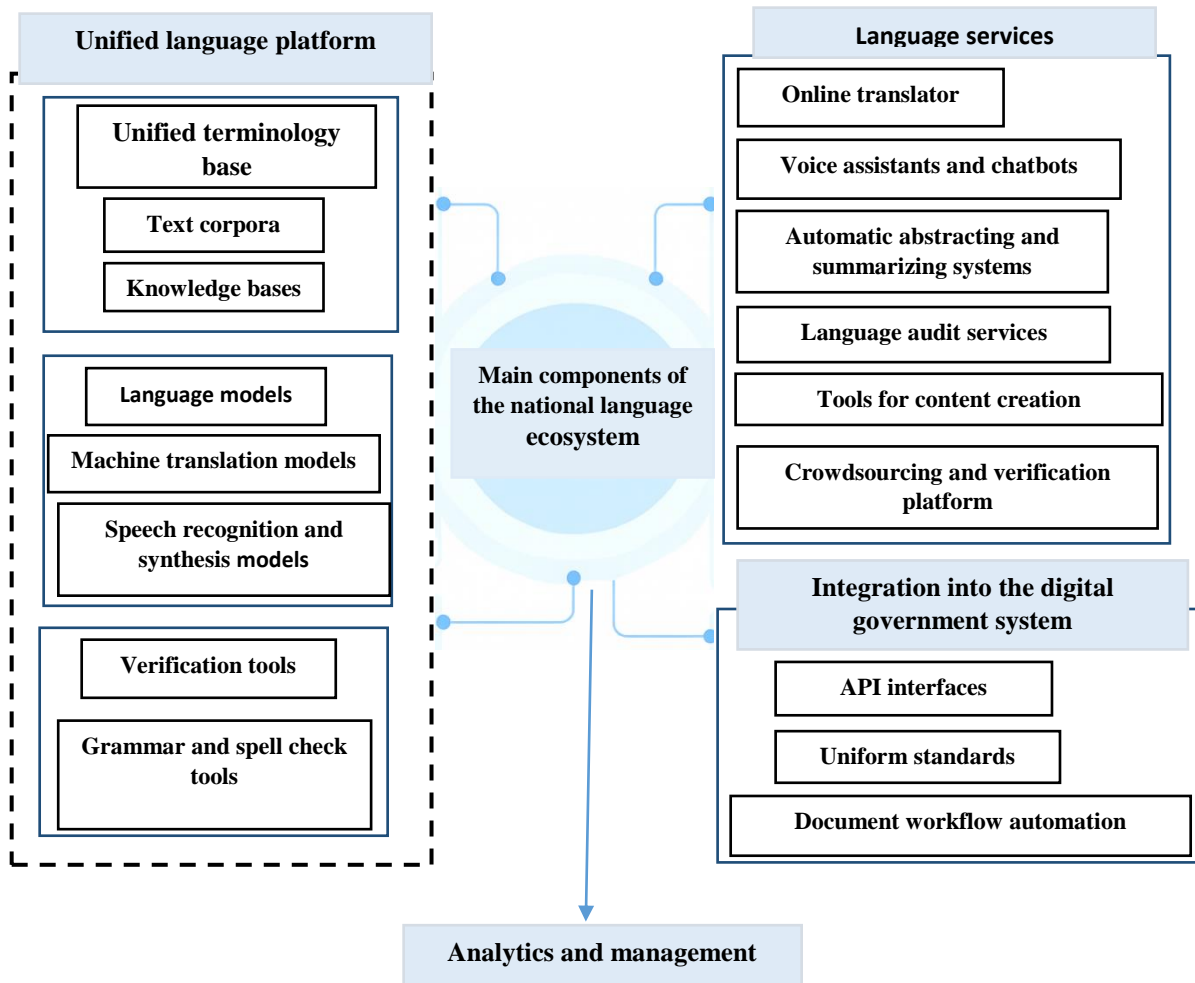


Fig. 3. Key components of the ecosystem

4. Discussion

This study analyzes the technological and functional sustainability of linguistic sovereignty in the virtual environment. The analysis demonstrates that the main threats to linguistic sovereignty (global linguistic hegemony, AI asymmetry, and content scarcity) are closely interconnected and can only be addressed through a comprehensive approach. Unlike the Westphalian model, modern linguistic sovereignty requires control over language not only within territory but also at the level of code, information (data), and services (AI tools). This makes national language policy an integral part of national security strategy.

The multi-level model of linguistic sovereignty in the digital environment we propose adequately addresses the challenges of the digital transformation era. In particular, one of the main contributions of the study is its emphasis on the level of technological sovereignty. Against the backdrop of asymmetries created by large language models,

the development of national language corpora and digital content systems is a survival strategy for small and medium-sized languages. Linguistic asymmetries in digital information systems are considered the most serious threat to linguistic sovereignty. Content sovereignty supports the functional viability of a language. In line with the digital cultural rights put forward by UNESCO, this level demonstrates that the promotion of high-quality digital content in the native language distracts users from using foreign-language resources and expands the functional scope of the national language in social media, education, and other digital platforms. In conclusion, the formation of a national language ecosystem and the unified and coordinated development of all its components are identified as the primary objective for protecting linguistic sovereignty in Azerbaijan in the era of digital transformation and globalization. Recommendations in this regard are provided.

5. Conclusion

The analysis shows that the global dominance of English, linguistic inequality on digital platforms, a lack of content for underresourced languages, and existing biases in artificial intelligence technologies create barriers to the full participation of national languages in the digital ecosystem. In the process of digitalization, expanding language resources, developing digital exchange technologies, improving digital literacy, and strengthening state regulatory support constitute the key pillars of linguistic sovereignty.

Developing a language ecosystem on a digital government platform is a complex issue aimed at ensuring effective communication between the state and citizens, as well as within government structures, taking into account multilingualism and the development of digital technologies. This isn't just a matter of translating interfaces, but rather creating a unified environment in which language becomes a tool for accessibility, inclusiveness, and the effectiveness of government services.

To ensure the digital sovereignty of national languages, the following measures are recommended:

- Developing strategic documents and action plans at the state level for the protection and development of the national language in the digital environment, integrating AI, digital data exchange technologies, and digital content into these strategies;
- Increasing targeted investments from the government and private sector in projects that integrate national languages with technologies such as NLP, ASD, OCD, speech synthesis, and machine translation;
- Implementing support programs aimed at creating and disseminating a variety of text, audio, video, and interactive materials in the national language;
- Expanding access to digital resources in the national language for all segments of society, improving digital literacy through educational and outreach programs;
- Digitizing dictionaries, grammar materials, folklore, and literary heritage, placing them in open databases, and making them accessible to researchers;
- Expanding the interface and content in national languages in collaboration with Facebook, TikTok, X (Twitter), and other international platforms;
- Forming a legislative framework for promoting digital content and services in the national language, creating monitoring and accountability mechanisms;
- Expanding cooperation with UNESCO, the European Union, ISO, and other international organizations, adapting international models to national realities;
- Expanding collaboration between specialists working in the field of language and technology, implementing interdisciplinary projects at universities and research institutes;
- Creating an open and accessible national language ecosystem that ensures linguistic sovereignty on the digital state platform and meets the needs of modern society.

These measures reinforce all components of the language ecosystem, enhancing the functionality of national languages in the digital space and ensuring linguistic sovereignty. The Azerbaijani language is also an integral part of this process. The development of a sustainable Azerbaijani language ecosystem on the state digital platform, in addition to ensuring its integration with modern technologies, is of strategic importance for the future development and promotion of the language.

Acknowledgements

We would like to express our sincere gratitude to Academician Rasim Alguliyev, the originator of this idea.

References

- Alguliyev, R. M., & Gurbanova, A. M. (2018). The conceptual foundations of national terminological information system. *IJ of Education and Management Engineering*, 8(4), 19-30. DOI: 10.5815/ijeme.2018.04.03
- Alguliyev, R.M. et al. (2025). Conceptual model of azerbaijani language ecosystem in the electronic government. *Comparative Cultural Studies: European and Latin American Perspectives*, (20), 1-24. <https://doi.org/10.46661/ccselap-11174>
- Alguliyev, R.M., Mahmudov, R.S. (2018). Language industry: opportunities, perspectives and problems. *Problems of Information society*, 1, 3-26. (in Azerbaijani) <https://doi.org/10.25045/jpis.v09.i1.01>
- Aliguliyev, R.M. et al. (2021). From cosmopolitan e-government to national e-government:

- prospects for creating Azerbaijani-language services in a virtual environment. *Problems of Information society*, №2, 3–25. <https://doi.org/10.25045/JPI.S.V12.I2.01>
- Anastasopoulos, A., Neubig, G. (2020). Should All Cross-Lingual Embeddings Be Aligned? Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 4514–4525. DOI: 10.18653/v1/2020.acl-main.766
- Barrinha, A., & Christou, G. (2022). Speaking sovereignty: the EU in the cyber domain. *European Security*, 31(3), 356-376. <https://doi.org/10.1080/09662839.2022.2102895>
- Bella, G. et al., (2023). Towards bridging the digital language divide. arXiv preprint arXiv:2307.13405. <https://doi.org/10.48550/arXiv.2307.13405>
- Bird, S. (2020). Decolonising speech and language technology. Proceedings of the 28th International Conference on Computational Linguistics, 3504–3519 DOI: 10.18653/v1/2020.coling-main.313
- Bradshaw, S., DeNardis, L. (2018). The politicization of the Internet’s Domain Name System: Implications for Internet security, universality, and freedom. *New media & society*, 20(1), 332-350. <https://doi.org/10.1177/14614448166629>
- Chen, J. et al., (2024). When large language models meet personalization: Perspectives of challenges and opportunities. *World Wide Web*, 27(4), 42. <https://link.springer.com/article/10.1007/s11280-024-01276-1>
- Council of Europe (2001). Council for Cultural Co-operation. Education Committee. Modern Languages Division. Common European framework of reference for languages: Learning, teaching, assessment. – Cambridge University Press, 278 p. <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4>
- Eckes, C. (2015). The reflexive relationship between internal and external sovereignty. *Irish J. European L.* 18, 33. https://www.researchgate.net/publication/272244952_The_Reflexive_Relationship_between_Internal_and_External_Sovereignty
- Floridi, L. (2020). The fight for digital sovereignty: What it is, and why it matters, especially for the EU. *Philosophy & technology*, 33(3), 369-378. <https://doi.org/10.1007/s13347-020-00423-6>
- Galla, C. K. (2016). Indigenous language revitalization, promotion, and education: Function of digital technology. *Computer Assisted Language Learning*, 29(7), 1137-1151. <http://dx.doi.org/10.1080/09588221.2016.1166137>
- Gazzola, M. (2016). Multilingual communication for whom? Language policy and fairness in the European Union. *European Union Politics*, 17(4), 546-569.
- Gurbanova A.M. (2015). On terminological threats to the Azerbaijani language in the context of globalization. *Problems of Information society*, №2, 87-95.
- Gurbanova, A. (2022, November). Problems and Prospects for Minority Languages in the Age of Industry 4.0. In the International Symposium on Computer Science, Digital Economy and Intelligent Systems (pp. 722-734). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-24475-9_59
- Haddow, B. et al., (2022). Survey of low-resource machine translation. *Computational Linguistics*, 48(3), 673-732. <https://aclanthology.org/2022.cl-3.6.pdf>
- Hayman And, P. A., Williams, J. (2006). Westphalian sovereignty: Rights, intervention, meaning and context. *Global society*, 20(4), 521-541.
- Hornberger, N. H. et al., (2018). Ethnography of language planning and policy. *Language Teaching*, 51(2), 152-186. <https://doi.org/10.1017/S0261444817000496>
- Hung, H. T. (2025). Exploring China’s cyber sovereignty concept and artificial intelligence governance model: a machine learning approach. *Journal of Computational Social Science*, 8(1), 24. <https://doi.org/10.1007/s42001-024-00346-8>
- Joshi, P. et al., (2020). The state and fate of linguistic diversity and inclusion in the NLP world. arXiv preprint arXiv:2004.09095. <https://doi.org/10.48550/arXiv.2004.09095>
- Kahiraba, K. M. K. (2025). The Impact of Globalization on Linguistic and Cultural Sovereignty in Mali. *Interdisciplinary Language and Culture Studies*. 3(1), 1-14. <https://doi.org/10.5281/zenodo.15756840>
- Kelly-Holmes, H. (2019). Multilingualism and technology: A review of developments in digital communication from monolingualism to idiolingualism. *Annual Review of Applied Linguistics*. 39, 24-39. <https://doi.org/10.1017/S0267190519000102>
- Kornai, A. (2013). Digital language death. *PLoS one*, 8(10), e77056.
- Kraus, Peter A. (2018). “Between minority protection and linguistic sovereignty”. *Revista de Llengua i*

- Dret, Journal of Language and Law, Issue 69, 6-17, DOI:10.2436/rld.i69.2018.3122
- Moshagen, S. N. et al., (2024). Indigenous language technology in the age of machine learning. *Acta Borealia*. 41(2), 102-116. <https://doi.org/10.1080/08003831.2024.2410124>
- Paris, R. (2020). The right to dominate: How old ideas about sovereignty pose new challenges for world order. *International Organization*, 74(3), 453-489.
- Phillipson, R. (2008). The linguistic imperialism of neoliberal empire. *Critical inquiry in language studies*. 5(1), 1-43. <https://doi.org/10.1080/15427580701696886>
- Rehm, G. et al., (2020, May). The European language technology landscape in 2020: Language-centric and human-centric AI for cross-cultural communication in multilingual Europe. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*. 3322-3332. <https://aclanthology.org/2020.lrec-1.407.pdf>
- Scannell, K. P. (2007). The Crúbadán project: Corpus building for under-resourced languages in Building and exploring Web corpora: Proceedings of the 3rd Web as Corpus Workshop. University Press of Louvain, vol. 4, 5-15. <https://kevinscannell.com/files/wac3.pdf>
- Siminyu, K. et al., (2021). AI4D--African Language Program. arXiv preprint arXiv:2104.02516. <https://arxiv.org/pdf/2104.02516>
- Tan, Y., Jehom, W. J. (2024). The Function of Digital Technology in Minority Language Preservation: The Case of the Gyalrong Tibetan Language. *Preservation, Digital Technology & Culture*, 53(3), 165-177.
- UNESCO. (2001). Universal Declaration on Cultural Diversity. Paris: UNESCO. <https://www.refworld.org/legal/resolution/unesco/2001/en/35901>
- UNESCO. (2023). Digital Initiatives for Indigenous Languages. Paris: UNESCO. https://www.defindia.org/wp-content/uploads/2024/12/Digital-Initiatives-for-Indigenous-Languages.pdf?utm_source=chatgpt.com

How to cite: Afruz Gurbanova, Mehriban Baghirova (2026). Linguistic sovereignty in virtual space: language ecosystem, problems and solutions. *Problems of Information Society*, 1, 103-116. <https://doi.org/10.25045/jpis.v17.i1.11>