International Journal on Culture, History, and Religion
https://ijchr.net | eISSN: 3028-1318

Received: May 12, 2025 | Accepted: June 28, 2025 | Published: September 30, 2025
Volume 7 Special Issue 1 | doi: https://doi.org/10.63931/ijchr.v7iSI1.195

Article

Intelligent Language Systems and Technical Linguistic Solutions in the Digital Environment

Iryna Shvetsova[1], Svitlana Lytvynska[2], Tetiana Davydova[3], Svitlana Romanchuk[4] and Olha Rusavska[5]

[1]Kherson State Maritime Academy, Kherson, Ukraine. [2]National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine. [3]National Academy of Security Service of Ukraine, Kyiv, Ukraine. [4]Ukrainian Institute of Art and Science, Bucha, Ukraine. [5]National University "Odesa Maritime Academy", Odesa, Ukraine.

Correspondence: phd.shvetsova@gmail.com

Abstract

In today's increasingly digital world, there is a growing demand for linguistic solutions that promote accessibility, ethics, and multilingualism in online environments. The challenges of language discrimination, digital inequality, and the endangerment of minority languages due to automated language processing are becoming increasingly urgent. This study aims to explore the role of applied linguistics in addressing these global language issues, with a particular focus on emerging trends in language engineering, framed within ethical and social contexts. The research adopts an interdisciplinary theoretical foundation, drawing from sociolinguistics, corpus linguistics, semiotics, cognitive science, and digital humanities. Through content analysis of relevant literature and data, the study typologizes ethical risks linked to the deployment of artificial intelligence-based language models and identifies key trends in applied linguistics, from socio-humanitarian interpretations of language to the construction of multilingual corpora for automated text processing. Findings highlight a tension between technocratic and sociocentric paradigms in current scientific discourse and underscore the absence of a unified ethical framework governing language technology. The practical value of this research lies in its proposed typology of ethical risks, which may inform digital inclusion policies, educational program design, and user interfaces for marginalized linguistic communities. Additionally, the study sets a foundation for future investigations into the intercultural adaptation of NLP systems, the creation of ethical protocols, and the development of normative approaches to linguistic diversity in the age of artificial intelligence.

Keywords: *applied linguistics, artificial intelligence, digital inequality, ethical risks, language discrimination, language engineering*

Suggested citation:

Shvetsova, I., Lytvynska, S., Davydova, T., Romanchuk, S., and Rusavska, O. (2025). Intelligent Language Systems and Technical Linguistic Solutions in the Digital Environment. *International Journal on Culture, History, and Religion, 7*(SI1), 297-313. https://doi.org/10.63931/ijchr.v7iSI1.195

Publisher's Note: IJCHR stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Introduction

In the 21st century, language is no longer just a subject studied at school – it is becoming a key tool for interacting with digital systems. The growth of digital communication, the development of multilingual chatbots, voice control interfaces, automated translators, and artificial intelligence language models is leading to a rethinking of the role of applied linguistics as a discipline capable not only of analyzing but also of constructing new language environments. At the same time, problems of language discrimination, digital inequality, the threat of extinction of small languages, and the lack of transparent ethical standards in the field of natural language processing are becoming more acute. For this, one needs to have a practical understanding of how applied linguistics can act as an area combining language, technology, and ethics. Equally important is the cultural perspective, as language is inherently tied to identity, heritage, and social meaning. In the digital era, where linguistic technologies shape everyday communication, media narratives, and access to information, applied linguistics becomes a powerful instrument of cultural transmission and intercultural mediation. Understanding how digital language tools affect cultural expression is essential for building inclusive and respectful communication systems.

Lately, both the social and technology areas of applied linguistics have been growing, helped by research on transformers, multilingual corpora, and language systems based on deep learning (Dodigovic & Jeaco, 2021; Zhang, 2023; Knight et al., 2021). More researchers are focusing on the ethics of AI within language, with concerns over discrimination, marginalization of minorities, and not granting everyone access to their language. Still, no general way to recognize common ethical problems in language engineering has been suggested among the many textbooks and articles. No one has clearly described how applied linguistics may tackle these global issues. The current literature addresses either the technical or ethical topic in a separate way, ignoring the fact that the mix of subjects is important. Threats coming from where language analytics, artificial intelligence, and social space meet have not yet been fully mapped. Research on how applied linguistics affects the development of an inclusive digital language setting is lacking.

This work aims to discuss how applied linguistics could address various global challenges, especially digital inequality, discrimination based on language, keeping narrow languages alive, and supporting all languages used in society. As a result, the following objectives are to describe the major approaches in modern applied linguistics, organize the main risks that language engineering research could present, and summarize interdisciplinary topics for new studies in the field.

Literature review

It is presented that applied linguistics now involves active interactions with technology, AI, and language ethics. The main scholars in this area are noting that applied linguistics should now be seen as a wide-ranging technology, not just as a field of linguistics. Widdowson (2019) emphasizes that language technologies should be developed within an interdisciplinary perspective, while Yaw et al. (2023) address the ethical considerations involved in research here. In their view, applied linguistics is a place where differences are listened to and understood. The use of language technologies in digital media is also noteworthy. Dodigovic and Jeaco (2021), Dmitrieva (2023), Lee (2025), and Narula and Chaudhary (2025) discuss how NLP, transformers, and multilingual models are helping to shape novel ways of communicating, mainly used in media, law, and education. The authors Dash (2021), Hunston (2022), and Knight et al. (2021) examine in detail why corpus linguistics and annotation help in training language models. They provide guidelines on how to put together corpora that result in producing effective training samples.

Many articles discuss the relationship between applied linguistics, linguistic engineering, and politics. They view applied linguistics as a way to represent the interests of minority languages in public policy. The studies of Malyuga (2024), Simanjuntak (2024), Zapf (2024), and Matthiessen et al. (2022) involve using language and cognition in professional and educational fields. In their books, Arau Ribeiro (2017) and Pennycook (2018) examine the history and theory of applied linguistics, seeing language as a complex idea system. The interpretation of data in digital format relies on the methodology analysis presented in Mulder (2020), Wang and Plonsky (2023) and Bubenhofer et al. (2018), among other authors.

Alduais et al. (2025) conducted a scientometric study, revealing that subjects including artificial intelligence, multilingualism, and digital humanities are gaining attention in applied linguistics. Štefčík (2025) also highlights the role of applied linguistics in helping the language industry transform translation, localization, and language consulting. All in all, the studies reveal that applied linguistics covers a wide range of themes: from examining corpora to considering ethical issues, from exploring brain processes to using artificial intelligence (Dodigovic & Jeaco, 2021; Hu, 2025; Kubanyiova & Creese, 2025; Widdowson, 2019). Because of this method, linguistics becomes a language science and supports technology, fairness, and communication worldwide.

Today's researchers often explore how to understand the concept of the linguistic subject. Pujolar and O'Rourke (2022) look at how individuals form their identity with language in multilingual societies, focusing on both social and

institutional factors. The same is true for Pennycook (2018), who believes language is a social practice and sees the scope of applied linguistics as including cultural criticism. Ziegler (2024) suggests that learning languages encourages individuals to become aware and involved citizens, especially when considering global developments. There are publications that explore how applied linguistics is affected by the rise of new technologies. As an illustration, Silvestri et al. (2022) show that a mix of digital humanities and corpus analysis can be applied to analyze sociocultural discourses. Yalçin (2023) discusses the way language models are changing in today's fast-changing digital and social environments. It is important to add that Zhang (2023) explores ethical language modeling for deep learning and automated decisions. Spasov (2023) also points out the risks of ignoring minority languages in language system creation with large language data.

Research by Papadopoulou and Tapio (2024) and Evangelia et al. (2024) raises the issue of access to language education for migrants and refugees, pointing to the shortcomings of current approaches to inclusive teaching. In the same vein, Wang and Plonsky (2023) criticize the methodological limitations of empirical research in applied linguistics, proposing to expand the methods of analysis in the field of language and technology. However, it should be noted that recent research, particularly by Hutson (2024) and Wetzel (2018), has already addressed strategies for preserving linguistic diversity in the digital age, offering scalable and culturally sensitive models. Therefore, this study approaches the issue not as entirely unexplored, but as requiring further integration into language system architectures. Thus, the existing scientific base outlines a wide range of approaches, where applied linguistics emerges both as a methodological system and as a socio-technical practice (Silvestri et al., 2022; Ziegler, 2024; Zhang, 2023; Papadopoulou & Tapio, 2024). However, despite the theoretical and practical diversity, some issues remain unresolved: first, the lack of a unified ethical framework for implementing language models in intercultural environments; second, the insufficient development of strategies for preserving linguistic diversity in the digital age.

Recent interdisciplinary contributions have expanded the scope of applied linguistics to include perspectives from ethics in AI (Hutson, 2024), cognitive modeling in digital ecosystems (Wetzel, 2018), and engineering strategies for linguistic preservation (Chen et al., 2024). These works underline the necessity of combining technological, cultural, and social viewpoints in constructing ethical language systems.

Methodology

A team of authors conducted the research in 2024–2025 based on an analysis of scientific literature, open digital resources, and relevant publications in the field of applied linguistics, language engineering, and natural language processing. The main method was qualitative content analysis of sources covering interdisciplinary aspects of applied linguistics. A comparative analysis of approaches to the use of language technologies in different contexts, such as educational, political, ethical, and technological, was applied. The method of systematization and typology of ethical risks based on secondary data from publications by leading researchers was also used. To form generalized conclusions, a bibliometric review based on the work of Alduais et al. (2025), Dodigovic and Jeaco (2021), Yaw et al. (2023), Kubanyiova and Creese (2025), and Widdowson (2019) was conducted, and for the construction of tables, the author's approach to classification based on meaningful grouping of topics was used. All sources were examined with regard to academic integrity and compliance with international standards. The literature was retrieved from scientometric databases including Scopus, Web of Science, SpringerLink, and Google Scholar. Selection criteria included the publication date (2017–2025), thematic relevance to applied linguistics, language technology, ethics, citation impact, and availability of full text.

Results

In today's digital society, applied linguistics has taken on a new meaning, transforming from a purely linguistic discipline into an interdisciplinary field that ensures effective human-computer interaction, artificial intelligence development, and language technologies. Many scientists agree that applied linguistics is a way to bring together language skills and technology to address real problems, such as automatic translation, speech processing, information collection and analysis, and computer text analysis. Researchers stress that applied linguistics analyzes language and also designs models for use in software (Yaw et al., 2023). The cognitive-communicative paradigm should be highlighted among today's approaches since it focuses on language as a social action in relation to technology. The approach helps to see applied linguistics as an element of humanities engineering. Here, Widdowson's (2019) viewpoint matters, since he points out that applied linguistics today is shifting to involve more areas of study and that language specialists should respond to the fluctuations in the digital area. This approach matters most in relation to the rise of large language models that work with natural language.

The idea of language technology has played an important role in what the world knows about applied linguistics today. According to Dodigovic and Jeaco (2021), language technologies help convert knowledge of languages into digital resources. Applied linguistics is used to create technology that allows humans and

machines to communicate and adapt to various languages and forms of communication across the Internet. In addition, applied linguistics helps ensure that language data processing is ethical by focusing on confidentiality, fairness of algorithms, and inclusivity (Kubanyiova & Creese, 2025). Therefore, due to the digital shift in society, applied linguistics is now focusing on both analyzing languages and making tools that encourage effective use of language in digital communications. For this transition to happen, people must have a strong grasp of language structure and of the current technologies available, which may lead to the development of new ways of researching and studying different fields together.

With AI and ML technology advancing fast today, linguistic engineering and speech processing are gaining importance in applied linguistics. These fields are actively transforming in line with the demands of the digital society, where natural language processing (NLP) is the basis for the creation of voice assistants, chatbots, automatic translation systems, speech recognition, emotion detection, and more. In this context, the main focus is on the development of deep neural networks, transformers (such as GPT and BERT models), approaches to multi- and cross-lingual modeling, as well as ethical and methodological aspects of artificial understanding of human language (Dodigovic & Jeaco, 2021; Kraus, 2023; Kubanyiova & Creese, 2025).

In view of the above, Table 1 illustrates the main current trends in the development of linguistic engineering and speech processing in light of innovations in AI and machine learning.

Table 1. Current trends in linguistic engineering and speech processing in the context of AI and machine learning

Direction of development	**Characteristics**	**Application areas**
Deep Learning	Creation of multi-level neural networks for modeling syntax, semantics, and pragmatics	Machine translation, text generation, chatbots
Transformer-based language models	BERT and GPT architectures that allow context processing at the level of large data arrays	Automatic text completion, dialogue systems
Multi- and cross-lingual technologies	Development of models capable of working with multiple languages or transferring knowledge from one language to another	Automatic translation, global information platforms
Speech recognition and synthesis	Algorithms for converting audio to text and vice versa using acoustic and linguistic models	Virtual assistants, voice control devices

Emotional speech processing	Integration of intonation, timbre, and rhythm analysis to detect emotions in the voice	Telemedicine, remote psychological diagnosis
Linguistic corpora and data annotation	Automated labeling of language resources for model training	Creating training samples for AI
Ethical aspects of NLP	Ensuring transparency, impartiality, and privacy	Regulation of AI solutions in education, law, and medicine

Source: created by the author based on Dodigovic and Jeaco (2021), Yaw et al. (2023), Kubanyiova and Creese (2025), Widdowson (2019)

In conclusion, it should be emphasized that linguistic engineering in the context of artificial intelligence is moving from formal language analysis to modeling it as a social and cognitive phenomenon capable of adaptation, learning, and interacting with humans in real time. Such areas as multilingualism, contextuality, emotionality, and ethics are becoming an integral part of the technological progress of language systems. Different vectors of development are playing a key role in changing how language is communicated.

Nowadays, language technologies have a major impact on the way education, media, and law function. As a result of being used in artificial intelligence and machine learning, they help to organize information and add new methods for communicating, interpreting, translating, and analyzing language data. In global approaches to this area, people combine knowledge about languages with computer models, making sure to consider ethical values, the use of many languages, and what users require. To illustrate, platforms for learning are using language models to help assess writing, adapt to each student's needs, and provide language support. In journalism, language technologies play a role in discussion analysis, detecting fake news, producing articles, and in the legal area, they help analyze court judgments and provide language knowledge (Dodigovic & Jeaco, 2021; Yaw et al., 2023; Kubanyiova & Creese, 2025).

Table 2 summarizes the main approaches to using language technologies in three main areas, as recommended by international experts and scientists.

Table 2. Integration of language technologies in education, media, and law: global approaches

Field	**Main technologies**	**Global practices and implementation models**	**Potential challenges**
Education	Automatic assessment, adaptive learning, machine translation, voice assistants	Platforms such as Duolingo, Grammarly, and Google Classroom use NLP for learning and feedback.	Algorithmic bias, privacy violations, and lack of cultural adaptation

Media	Text generation (NLG), discourse analysis, fake detection, automatic subtitling	OpenAI Codex, Meta LLaMA, BBC automate news translation; NYT uses AI for content analysis	Ethics, audience manipulation, copyright
Law	Analysis of legal documents, legal expertise, search for precedents, AI court systems	NLP-based systems in LexisNexis, CaseText, smart contract projects based on GPT	Legal liability, confidentiality, risk of errors

Source: created by the author based on Dodigovic and Jeaco (2021), Yaw et al. (2023), Kubanyiova and Creese (2025), and Widdowson (2019)

Thus, the integration of language technologies into education, media, and law demonstrates the powerful potential for transforming communication, analytics, and access to knowledge on a global scale. At the same time, global approaches emphasize the need for interdisciplinary balance: a combination of technological efficiency, linguistic accuracy, and social and ethical responsibility. This combination enables the sustainable development of intelligent systems that work with human language.

In a global context, linguistic engineering and speech processing have become hotbeds of high-tech competition between leading research centers and tech giants such as OpenAI, Google AI, Meta AI, Amazon, and Microsoft Research. The vector of development is focused not only on the quality of speech recognition or text generation, but also on scalability, multilingualism, the ability to learn without a teacher, and the ethical sustainability of models. Since 2018, there has been exponential growth in the volume of training corpora, the number of parameters in language models, and the power of their application in open ecosystems. In this dynamic environment, leading players are setting benchmarks for both the scientific and applied communities.

The research methodology consists of a comparative analysis of open data on key linguistic platforms and models that implement the principles of language engineering. The sources are both technical specifications of models (e.g., GPT-3, BERT, Whisper, LLaMA) and accompanying scientific analytics. The following indicators are compared: launch year, training data volume (in terabytes), number of languages, number of parameters in models (billion), and type of openness (open-source or proprietary). All data is presented in Table 3 for further visualization in graphical form.

Table 3. Comparative characteristics of leading language technologies (2018–2024)

Platform/ Model	Year of launch	Training data volume (TB)	Number of parameters (billion)	Number of languages	Access type

BERT (Google AI)	2018	~16	0,34	104	Open-source
GPT-3 (OpenAI)	2020	~45	175	1 (English)	Proprietary
Whisper (OpenAI)	2022	~680	1,55	97	Open-source
LLaMA 2 (Meta AI)	2023	~200	65	20+	Open-source
Gemini (Google DeepMind)	2024	~800	280	30+	Partially open

Source: created by the author based on Dodigovic and Jeaco (2021), Yaw et al. (2023), Kubanyiova and Creese (2025), and Widdowson (2019)

As can be seen from the table, there is a stable and rapidly growing trend in the scaling of language models. For example, in five years, the number of parameters increased from 0.34 billion (BERT, 2018) to 280 billion (Gemini, 2024), which is an 823-fold increase. The volume of training corpora has grown even more impressively – from 16 TB to over 800 TB, i.e., 50 times. There has also been a noticeable expansion of language coverage: from 1 to over 100 languages (BERT, Whisper), indicating a shift from English-centric models to global multilingual systems. Interestingly, despite the growth of proprietary platforms, most research models (Whisper, LLaMA) remain open-source, which promotes transparency and academic rethinking of language technologies. In addition, models such as Whisper specialize in speech recognition and have unique corpora that include audio, which distinguishes them from predominantly text-centric models such as GPT-3.

The development of linguistic engineering systems is accompanied by a rapid increase in the amount of training data used to train language models. Figure 1 shows the change in the size of training corpora (in terabytes) for leading language technologies between 2018 and 2024.

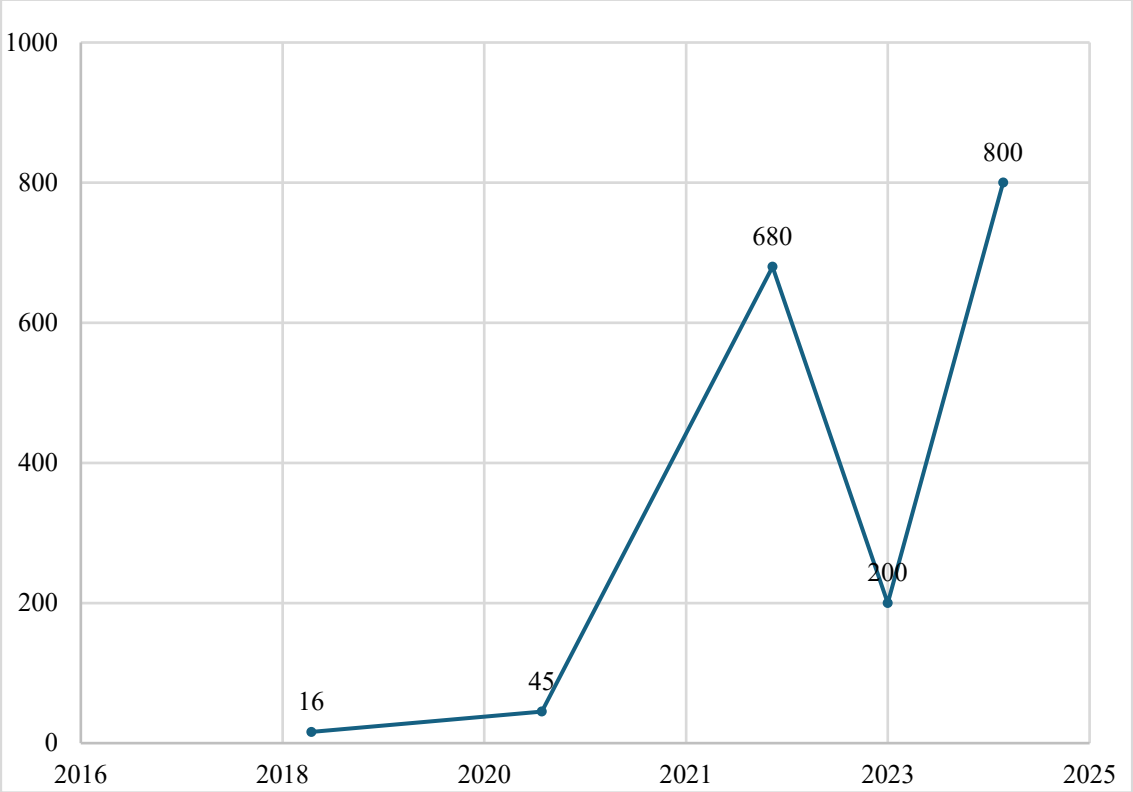

Figure 1. Dynamics of training data volumes (TB) in leading language models (2018–2024)
Source: created by the author based on Dodigovic and Jeaco (2021), Yaw et al. (2023), Kubanyiova and Creese (2025), and Widdowson (2019)

In 2018, the BERT model (Google AI) used a 16 TB training corpus, which became the baseline for subsequent generations of language models. By 2020, this volume had increased to 45 TB in GPT-3 (OpenAI), representing an increase of 29 TB, or approximately 181% in two years. The most dramatic growth was observed in 2022 with the release of Whisper from OpenAI, which already used 680 TB – almost 15 times more than GPT-3. However, in 2023, there was a significant decline to 200 TB in LLaMA 2 (Meta AI), which can be explained by training optimization or a shift in focus to model architectural efficiency. At the same time, in 2024, Gemini (Google DeepMind) again reached a new high of 800 TB, exceeding the previous record by 120 TB (or 17.6%). Overall, between 2018 and 2024, the volume of training data grew 50-fold, indicating a steady exponential growth in the development of language models and a high rate of technological improvement in the field of linguistic engineering.

Thus, the key areas of language technology development are concentrated around five axes: model scalability, multilingualism, open source, specialization (recognition, generation, translation), and ethical engineering. The experience of platforms such as OpenAI, Google AI, and Meta AI demonstrates a new era of language modeling – not only technological, but also value-based, where accessibility,

inclusivity, and safety for users from different linguistic and cultural backgrounds are priorities.

In a globalized digital world where language interaction takes place through technological interfaces, applied linguistics plays a key role in ensuring equitable access to information, education, and social services. The application of its tools not only expands the functionality of language technologies but also addresses fundamental global challenges such as digital inequality, language discrimination, the decline of small languages, and the inaccessibility of educational content for linguistic minorities. Applied linguists believe that technology makes it possible for language to be used as a tool for bringing people together, not separating them (Kubanyiova & Creese, 2025; Yaw et al., 2023; Galang-Pereña, F., 2024). It applies in situations with multilingual AI, creating programs for different regions and maintaining language diversity.

Table 4 lists international challenges that applied linguistics can help overcome, along with several examples of useful tools and strategies.

Table 4. The potential of applied linguistics in addressing global language and communication challenges

Global problem	**The role of applied linguistics**	**Examples of technologies/strategies**
Digital inequality	Development of interfaces in local languages; reduction of language barriers	Automatic translation, voice assistants for illiterate users (e.g., Google Speech-to-Text)
Language discrimination	Creating language-inclusive models, preventing algorithmic bias	Algorithms for checking bias, multicultural annotation of corpora
Decline of small languages	Digital recording, corpus creation, automated annotation, and generation	Language preservation programs (such as UNESCO Endangered Languages), Mozilla Common Voice
Inaccessibility of education for linguistic minorities	Development of educational content in native languages, access via mobile applications	Localization of platforms (Coursera, Khan Academy), machine translation in educational environments
Inequality in legal access	Language support in court and administrative proceedings	NLP systems for translating court decisions, automated legal advice (e.g., DoNotPay)
AI bias towards dominant languages	Multilingual models and corpora including small and regional languages	LLaMA, NLLB (No Language Left Behind) from Meta

Source: created by the author based on: Kubanyiova and Creese (2025), Yaw et al. (2023), Dodigovic and Jeaco (2021), and Widdowson (2019)

So, applied linguistics can be important not only in technology but also in society. These tools help address problems of unequal access to knowledge, services, and legal documents. When researching digital transformation, experts consider ethical questions, protect various cultures and languages, and focus on making development sustainable.

AI language models are being used, leading to several ethical issues that impact applied linguistics. These models also help shape the way people communicate, understand meaning, form opinions, and discuss matters in public. As a result, a list of ethical risks tied to language technologies should be prepared for use in education, journalism, public administration, and justice.

The first example of an ethical risk is when people are reinforced in their biases. Language models are built using huge texts that include historical and cultural inequality examples. Therefore, they might unthinkingly spread stereotypes about race, gender, religion, or language. It becomes more hazardous when AI is used to make choices in legal or medical fields. To prevent this, it is necessary to apply ethical auditing methods to corpora, ensure transparent documentation of sources, and verify annotation practices. The second risk is a lack of inclusivity due to a predominant focus on English or other dominant languages. This situation leads to small or indigenous languages and languages with complex morphology being left out of digital representation. This contributes to the digital marginalization of language communities. This can be prevented by developing multilingual models, such as Meta's NLLB (No Language Left Behind), or by creating national language preservation initiatives focused on linguistic justice. The third type of ethical risk lies in the opacity of how models work. Most modern models, especially commercial ones, are closed to analysis, and their algorithmic decisions remain incomprehensible even to developers (the so-called black box effect). This makes it impossible for society to control its decisions and prevents the establishment of cause-and-effect relationships in the event of errors. One way to overcome this problem is to promote explainable AI policies, which require algorithms to be described in a way that is understandable to the end user. The risk of involuntary user participation in AI training requires special attention.

Language models sometimes use open communications (e.g., forums, personal blogs) without the authors' consent. This poses an ethical threat to privacy and contradicts the basic principles of information security. This issue should be addressed by introducing a transparent data collection system where the participation of database authors is voluntary and the data is depersonalized.

Thus, ethical risks associated with AI language models encompass both technical and deeply social aspects, ranging from the reproduction of discriminatory patterns to the invisible exclusion of linguistic minorities. The principles of *transparency, inclusivity, and non-discrimination* must not be merely declarative, but embedded at the level of language system architecture, training corpora, interfaces, and evaluation procedures. Applied linguistics acts as a critical bridge between linguistic expertise and digital ethics in this process.

Discussions

The results of the study confirm the hypothesis that applied linguistics, being at the intersection of humanities and technical knowledge, is a key tool for addressing the global challenges of the digital age, including language discrimination, digital inequality, and the loss of linguistic diversity. From a review of the literature, it appears that half of the researchers support changing the way applied linguistics operates in terms of ethics (Kubanyiova & Creese, 2025; Yaw et al., 2023), while the other half focuses on the technology the field uses (Dodigovic & Jeaco, 2021; Zhang, 2023). Consequently, people can discuss how to balance fulfilling social duties and advancing in language engineering. A number of researchers (Widdowson, 2019; Pennycook, 2018) believe that applied linguistics should no longer simply focus on language analysis and should concentrate on social action. Similarly, Knight et al. (2021) and Hunston (2022) consider applied linguistics as a field that uses technology and statistics to analyze language. This contrast highlights the split in approaches to the nature of the object of study: on the one hand, language as a tool for social interaction, and on the other, as data for computation.

The present study attempts to integrate these approaches, which has created a typology of ethical risks and demonstrated the potential of applied linguistics in both the social sciences and humanities and engineering. The typology of ethical risks identified in the study includes the following categories: (1) Bias reinforcement – the reproduction of discriminatory stereotypes through training data; (2) Exclusion through monolingual focus, marginalization of speakers of less represented languages; (3) Opacity of algorithmic decision-making, the lack of transparency in model behavior and outputs; (4) Unconsented data usage, use of user-generated content without consent for training language models. These risks are summarized in Table X and represent core areas for the ethical evaluation of AI-based linguistic systems. Unlike Silvestri et al. (2022) and Leoni de León (2023), who emphasize the application of digital humanities to the analysis of language policy, the authors of this study focus on the need for a joint regulatory approach to multilingual NLP. At the

same time, the results of our analysis confirm the views of Spasov (2023) and Evangelia et al. (2024) regarding threats to linguistic minorities in digital ecosystems, but the authors believe that solving this problem requires not only language localization but also the expansion of inclusive practices at the model architecture level. Different approaches can also be seen in the issue of ethical responsibility. Zhang (2023) and Hu (2025) propose a concept of language system transparency based on semiotics and explainable AI, while Pujolar and O'Rourke (2022) emphasize the need for political representation of speakers. Our analysis has limitations: it is based only on secondary sources and does not include empirical surveys of users or language model developers. This creates a basis for future research. At the same time, there are limitations to the analysis: it is based solely on secondary sources and does not include empirical surveys of users or language model developers. This provides a basis for future research, particularly in the direction of analyzing the interaction of language technology with marginalized groups, as well as the development of global ethical standards for language engineering.

Thus, modern applied linguistics requires further research to develop an integrative model that takes into account both technological innovations and sociocultural responsibility. The need for interdisciplinary consensus, particularly between philologists, engineers, and lawyers, emerges as a key condition for the sustainable development of language technologies in the global digital environment.

Attention to the cultural dimension of applied linguistics is particularly significant in the context of digital transformation. Language technologies increasingly influence how people communicate and how cultural values, traditions, and collective identities are represented and transmitted. The integration of linguistic systems into education, media, and public administration must therefore account for ethical, technological, and cultural factors. Applied linguistics enables language model design that reflects local cultures' diversity and specificity through methods such as cultural semiotics, discourse analysis, and localization strategies.

Conclusions

The analysis showed that applied linguistics has already gone beyond the traditional linguistic paradigm and works as a powerful tool for interpreting, transforming, and regulating the linguistic space in a digital society. The main innovation of this study lies in its attempt to integrate the humanities and technical vectors of the field's development by focusing on ethical risks, inclusion, and transparency of language models. Unlike many existing studies that fragmentarily consider technological or social aspects, the research offers a typological approach to

analyzing the potential of applied linguistics as an interdisciplinary resource. The practical significance of the results lies in their potential use as a methodological basis for the development of language policy, ethical protocols for NLP systems, and educational strategies for the digital age. Among the main limitations, it is worth noting the lack of empirical data from representatives of language communities and AI developers, which limits the depth of applied analysis. At the same time, the results of the study open up prospects for interdisciplinary research involving cognitive scientists, lawyers, ethical engineers, and linguists. In the future, it would be useful to study the mechanisms of linguistic inclusion in automated communication and develop normative models for adapting language systems to multi-ethnic and multilingual contexts. Of particular relevance is the creation of international ethical protocols for the architecture of language models to ensure linguistic fairness and transparency of decisions made with the participation of artificial intelligence.

Funding

This research received no external funding.

Conflicts of Interests

The authors declare no conflict of interest.

References

[1] Alduais, A., Yassin, A. A., & Allegretta, S. (2025). Computational linguistics: A scientometric review. *Quality & Quantity*. Advance online publication. https://doi.org/10.1007/s11135-025-02138-2

[2] Arau Ribeiro, M. del C. (2017). Historical linguistics, linguistics, and applied linguistics: A study inspired by trees. *Odisea, (12),* 218. https://doi.org/10.25115/odisea.v0i12.218

[3] Bubenhofer, N., Rothenhäusler, K., Affolter, K., & Pajovic, D. (2018). The linguistic construction of world: An example of visual analysis and methodological challenges. In: N. Bubenhofer (Ed.), *Quantifying approaches to discourse for social scientists.* (pp. 251–284). Springer. https://doi.org/10.1007/978-3-319-97370-8_9

[4] Dash, N. S. (2021). Principles and rules of part-of-speech annotation. In: *Language Corpora Annotation and Processing.* (pp. 25–43). Springer. https://doi.org/10.1007/978-981-16-2960-0_2

[5] Dmitrieva, A. (2023). The role of language technology in accessible communication research. In: *Emerging Fields in Easy Language and Accessible Communication Research.* (pp. 319–337). Springer. https://doi.org/10.57088/978-3-7329-9026-9_12

[6] Dodigovic, M., & Jeaco, S. (2021). Technology in applied linguistics. *International Journal of TESOL Studies, 3*(2), 1–4. https://doi.org/10.5281/zenodo.5137155

[7] Galang-Pereña, F. (2024). Language: The Soul of Culture and Heart of Identity. *International Journal on Culture, History, and Religion*, 2(1), 21–26. https://doi.org/10.63931/ijchr.v2i1.48

[8] Hu, Y. (2025). Transparent languages. In: *Semiotics of the Image.* (pp. 51–81). Springer. https://doi.org/10.1007/978-981-96-4592-3_3

[9] Hunston, S. (2022). Designing a corpus. In: *Corpora in applied linguistics.* (pp. 18–46). Cambridge University Press. https://doi.org/10.1017/9781108616218.002

[10] Knight, D., Morris, S., Arman, L., Needs, J., & Rees, M. (2021). Understanding the language context. In: *Building a National Corpus.* (pp. 1–22). Springer. https://doi.org/10.1007/978-3-030-81858-6_1

[11] Kraus, P. A. (2023). Language and the politics of recognition. In: *Language and Recognition: Multilingualism and the Politics of Identity.* (pp. 45–66). Springer. https://doi.org/10.1007/978-3-031-15011-1_3

[12] Kubanyiova, M., & Creese, A. (2025). Introduction: Applied linguistics, ethics and aesthetics of encountering the other. *Applied Linguistics Review, 16*(2), 589–607. https://doi.org/10.1515/applirev-2024-0083

[13] Lee, R. (2025). Natural language processing. In: *Natural Language Processing.* (pp. 3–18). Springer. https://doi.org/10.1007/978-981-96-3208-4_1

[14] Leoni de León, J. A. (2023). The relationship between language policy and planning, theoretical linguistics and natural language processing. In: *Epistemological and Theoretical Foundations in Language Policy and Planning.* (pp. 101–127). Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-22315-0_5

[15] Malyuga, E. N. (2024). The linguistic-cultural dimension of corporate communication. In: *The Language of Corporate Communication.* (pp. 83–151). Springer. https://doi.org/10.1007/978-3-031-58905-8_2

[16] Matthiessen, C. M. I. M., Wang, B., Ma, Y., & Mwinlaaru, I. N. (2022). Cognition in systemic functional linguistics. In: *Systemic Functional Insights on Language and Linguistics.* (pp. 147–195). Springer. https://doi.org/10.1007/978-981-16-8713-6_6

[17] Mulder, G. (2020). The new statistics for applied linguistics. *Dutch Journal of Applied Linguistics, 9*(1), 79–96. https://doi.org/10.1075/dujal.19019.mul

[18] Narula, R., & Chaudhary, P. (2025). A comprehensive review on detection of hate speech for multilingual data. *Social Network Analysis and Mining, 14*, 244. https://doi.org/10.1007/s13278-024-01401-y

[19] Pennycook, A. (2018). Applied linguistics as epistemic assemblage. *AILA Review, 31*(1), 113–134. https://doi.org/10.1075/aila.00015.pen

[20] Pujolar, J., & O'Rourke, B. (2022). Theorizing the speaker and speakerness in applied linguistics. *Journal of Applied Linguistics and Professional Practice, 16*(2). https://doi.org/10.1558/jalpp.22760

[21] Simanjuntak, R. R. (2024). *Claiming importance of research: A corpus linguistics analysis on Indonesian students 'research papers*. In: *Applied Linguistics in the Indonesian Context.* (pp. 155–179). Springer. https://doi.org/10.1007/978-981-97-2336-2_9

[22] Štefčík, J. (2025). Language industry. In: *Multidisciplinary Insights into Translation Studies.* (pp. 119–178). Springer. https://doi.org/10.1007/978-3-031-87205-1_3

[23] Wang, C. S., & Plonsky, L. (2023). A typology of secondary research in applied linguistics. *Applied Linguistics Review, 15*(4), 1569–1594. https://doi.org/10.1515/applirev-2022-0189

[24] Widdowson, H. G. (2019). Disciplinarity and disparity in applied linguistics. In: *BAAL 2018: Taking Risks in Applied Linguistics.* (pp. 33–49). White Rose University Press. https://doi.org/10.22599/baal1.c

[25] Yaw, K., Plonsky, L., Larsson, T., Sterling, S., & Kytö, M. (2023). Research ethics in applied linguistics. *Language Teaching, 56*(4), 478–494. https://doi.org/10.1017/s0261444823000010

[26] Zapf, M. (2024). *Sprache und Kognition*. In: *Gender, Sprache, Kognition.* (pp. 69–137). Springer. https://doi.org/10.1007/978-3-662-69024-6_3