INTERNATIONAL MULTI DISCIPLINARY JOURNAL FOR RESEARCH & DEVELOPMENT
eISSN 2394-6334 Impact factor: 8.854 Volume 13, issue 05 (2026)

THE ROLE OF ARTIFICIAL INTELLIGENCE IN PHILOLOGICAL RESEARCH

Shukurova Bahor Boltayevna
Senior Lecturer, Department of Uzbek Language and Literature,
Termiz State University of Engineering and Agrotechnologies,
PhD in Philological Sciences
https://doi.org/10.5281/zenodo.20054670

Abstract: This article examines the role of artificial intelligence in modern philological research. In recent years, artificial intelligence technologies, especially natural language processing, machine translation systems, corpus linguistics tools and large language models, have become increasingly important in the study of language, literature, text, discourse and cultural communication. Philology, as a field that traditionally relies on textual analysis, historical-comparative methods, interpretation and linguistic description, is now entering a new methodological stage under the influence of digital technologies. Artificial intelligence enables researchers to process large volumes of textual data, identify lexical and semantic patterns, compare linguistic units across languages, analyze stylistic features and create digital corpora for further investigation. At the same time, the use of artificial intelligence in philology requires critical thinking, methodological caution and ethical responsibility, because machine-generated results may contain inaccuracies, cultural bias or contextual limitations.

Keywords: artificial intelligence, philology, natural language processing, corpus linguistics, digital humanities, linguistic analysis, text interpretation, machine translation, large language models, Uzbek language.

Introduction. The rapid development of artificial intelligence has significantly influenced almost all areas of scientific research, including the humanities. Philology, which studies language, literature, written monuments, speech culture and textual meanings, is one of the disciplines that can benefit greatly from artificial intelligence technologies. Traditionally, philological research has been based on close reading, linguistic comparison, historical interpretation, stylistic analysis and semantic observation. These methods remain essential today; however, the emergence of digital tools has created new opportunities for expanding the scale, speed and accuracy of research.

Artificial intelligence is increasingly used in education and research as a human-centred technological tool, and UNESCO emphasizes that generative AI should be applied with attention to human agency, ethical use and institutional responsibility. In philological research, this means that AI can assist in collecting, classifying and interpreting linguistic material, but the final scientific judgement must remain with the researcher.

The relevance of this topic is determined by several factors. First, modern philology deals with large amounts of textual information: literary works, historical manuscripts, dictionaries, dialect materials, media texts, oral speech transcripts and digital communication data. Second, the development of corpus linguistics and digital humanities has changed the way researchers approach language material. Third, artificial intelligence tools now make it possible to analyze linguistic phenomena not only manually, but also through automated and semi-automated methods. Stanford's 2025 AI Index describes artificial intelligence as one of the most transformative technologies of the present era and highlights its growing influence on research, economy and society.

Therefore, the study of artificial intelligence in philology is not only a technological issue, but also a methodological and theoretical problem. It requires a clear understanding of how AI

can be integrated into linguistic and literary research without weakening the scientific depth, cultural sensitivity and interpretive nature of philology.

Literature Review. The relationship between artificial intelligence and philology is closely connected with the development of natural language processing, corpus linguistics and digital humanities. Natural language processing, usually abbreviated as NLP, is a branch of artificial intelligence that focuses on the interaction between computers and human language. Through NLP, computers can recognize, classify, translate, summarize and generate human language.

In digital humanities, language is considered one of the main objects of computational analysis. The NLP4DH research community describes its field as the application of natural language processing techniques to digital humanities research, including text analysis, text processing and error analysis of NLP systems from humanities perspectives. This approach is especially important for philology, because philological research often requires both quantitative observation and qualitative interpretation.

Recent studies also show that large language models can support qualitative data analysis by assisting with document classification, information extraction, span classification and text generation. These functions are directly relevant to philological research, where scholars often work with large collections of texts and need to identify recurring lexical, grammatical, semantic or stylistic patterns.

Another important direction is the use of artificial intelligence in the analysis of multilingual and historical documents. Research on large language model integration in digital humanities notes that the growth of multilingual historical data creates a need for effective, efficient and explainable tools for extracting information from such materials. This is especially significant for Uzbek philology, because the Uzbek language has a rich written tradition connected with Arabic script, Latin script, Cyrillic script and modern Latin-based orthography.

Thus, the existing literature demonstrates that artificial intelligence is not external to philology. On the contrary, it is becoming part of a broader methodological transformation in humanities research.

Research Methodology. This article uses a theoretical-analytical research method. The study is based on the analysis of scientific approaches to artificial intelligence, natural language processing, corpus linguistics and digital humanities. The methodological basis of the article includes descriptive analysis, comparative analysis, systematic interpretation and conceptual generalization.

The descriptive method is used to explain the basic functions of artificial intelligence in philological research. The comparative method helps to identify the difference between traditional philological analysis and AI-assisted analysis. The systematic method allows artificial intelligence to be understood not as an isolated technical instrument, but as part of a broader scientific research system. The conceptual approach is used to define the methodological significance of AI in the study of language and literature.

The article does not treat artificial intelligence as a replacement for human intellectual activity. Instead, it views AI as a research-supporting mechanism that can strengthen the philologist's ability to collect, classify, compare and interpret textual data.

Artificial intelligence offers several important possibilities for philological studies. One of the most important is the ability to process large-scale textual data. In traditional research, a philologist may analyze a limited number of texts manually. This method is deep and interpretive, but it requires much time. AI-based tools, by contrast, can process thousands or even millions of words in a short period of time. This allows the researcher to identify patterns that may remain invisible during manual reading.

For example, artificial intelligence can help determine the frequency of certain words, phrases, grammatical constructions or stylistic devices in a literary text. It can compare the language of different authors, periods or genres. It can also identify semantic fields, metaphorical patterns, discourse markers and syntactic tendencies. Such possibilities are especially useful in corpus linguistics, stylistics, lexicography, translation studies and comparative literature.

In linguistic research, artificial intelligence can be used to study phonetics, morphology, syntax, semantics and pragmatics. In phonetic research, AI-based speech recognition systems can help transcribe oral speech and analyze pronunciation features. In morphology, AI can identify word forms, roots, affixes and grammatical categories. In syntax, it can analyze sentence structure and dependency relations. In semantics, it can detect meaning relations, synonymy, antonymy, polysemy and contextual usage. In pragmatics, it can help analyze communicative intention, politeness strategies, speech acts and discourse organization.

In literary studies, artificial intelligence can assist with authorship attribution, genre classification, motif detection and stylistic comparison. For instance, if a researcher studies the style of a particular writer, AI can help identify frequently used lexical units, sentence patterns, figurative expressions and narrative structures. However, the interpretation of these results still requires philological knowledge. A machine can show what is repeated, but the scholar must explain why it is meaningful.

Corpus linguistics is one of the areas where artificial intelligence has become particularly useful. A corpus is a large, structured collection of texts used for linguistic analysis. Modern corpora may include literary texts, newspaper articles, academic writings, oral speech transcripts, dialect materials and social media texts.

Artificial intelligence can help create, annotate and analyze corpora. Corpus annotation means adding linguistic information to texts, such as part of speech, lemma, syntactic function or semantic category. Manual annotation is very time-consuming, while AI can significantly accelerate this process. As a result, researchers can conduct more complex and extensive analyses.

For Uzbek philology, corpus linguistics has special importance. The Uzbek language has undergone several alphabet changes, and its vocabulary includes Turkic, Arabic, Persian, Russian and international elements. Artificial intelligence can help study these layers more systematically. It can assist in preparing electronic dictionaries, identifying historical lexical changes, comparing dialectal variants and analyzing modern usage.

AI-based corpus analysis can also support educational research. For example, it can help determine which words are most frequently used in school textbooks, which grammatical structures create difficulties for learners, or how students use language in written assignments. Such data can be useful for improving language teaching methodology. Machine translation is one of the most visible applications of artificial intelligence in philology. Modern translation systems use large datasets and neural models to translate texts from one language into another. These tools can be useful for preliminary translation, terminology comparison and multilingual communication.

However, philological translation is not merely the replacement of words from one language with words from another. Translation requires cultural understanding, stylistic sensitivity, contextual interpretation and knowledge of national mentality. Therefore, artificial intelligence can support the translator, but it cannot fully replace the human translator.

This issue is especially relevant for the Uzbek language. Uzbek has specific grammatical structures, word order patterns, cultural expressions, idioms and speech etiquette forms. Machine translation systems may incorrectly translate idiomatic expressions, metaphorical meanings or

culturally marked units. For example, proverbs, phraseological units and poetic images often require interpretation rather than literal translation.

Thus, artificial intelligence should be used in translation studies as an auxiliary tool. It can provide draft versions, identify terminology equivalents and compare translation variants. But the final translation must be edited and evaluated by a qualified specialist.

Literary text analysis is one of the most complex areas of philology because literary meaning is often implicit, symbolic, emotional and culturally conditioned. Artificial intelligence can help identify formal features of literary texts, but it cannot fully understand aesthetic value in the same way as a human reader.

AI can be useful in analyzing the vocabulary of literary works, identifying themes, comparing characters' speech, detecting emotional tone and studying narrative structure. For example, sentiment analysis may show whether a passage contains positive, negative or neutral emotional colouring. Topic modelling may help identify dominant themes in a large collection of texts. Stylometric analysis may reveal similarities or differences between authors.

Nevertheless, literary analysis cannot be reduced to statistics. A word may have different meanings depending on context, irony, symbolism or historical background. Therefore, AI-generated data should be interpreted carefully. The philologist must connect quantitative findings with literary theory, cultural context and textual interpretation.

In this sense, artificial intelligence can be compared to a microscope in literary research. It allows the scholar to see details more clearly, but it does not explain the whole meaning by itself. The explanation remains the task of the researcher.

The application of artificial intelligence in Uzbek philology has significant scientific potential. Uzbek philological research includes the study of literary language, dialects, historical texts, oral folklore, terminology, phraseology, lexicography and translation. Each of these areas can benefit from AI-based methods.

First, AI can help develop digital corpora of the Uzbek language. Such corpora can include classical literature, modern prose, poetry, scientific texts, journalistic materials and spoken language. Second, AI can assist in creating electronic dictionaries, including explanatory, bilingual, terminological and phraseological dictionaries. Third, it can support the study of dialects by comparing regional lexical and phonetic features.

Artificial intelligence may also be useful for the study of classical Uzbek literature. Many historical texts exist in manuscripts or old printed editions. AI-based optical character recognition systems can help digitize these texts, although the recognition of Arabic-script or old orthographic materials remains a difficult task. After digitization, researchers can analyze vocabulary, imagery, stylistic devices and intertextual connections more systematically.

Another promising direction is the creation of Uzbek-language AI tools. Many existing systems work better for English and other widely represented languages because they have more training data. Uzbek, like many other languages, needs more digital resources, annotated corpora and high-quality datasets. Therefore, philologists, linguists and computer scientists should cooperate in developing language technologies for Uzbek.

The use of artificial intelligence in philology has several advantages.

First, it increases research efficiency. AI can process large textual materials much faster than manual methods. This saves time and allows researchers to work with broader data.

Second, it improves the objectivity of certain types of analysis. For example, word frequency, collocation patterns and syntactic structures can be counted automatically. This helps support theoretical conclusions with empirical evidence.

Third, it expands interdisciplinary research. Philology can cooperate with computer science, data science, cognitive linguistics, cultural studies and education. Such cooperation creates new research problems and methods.

Fourth, artificial intelligence can help preserve cultural heritage. Manuscripts, folklore materials, old newspapers and rare books can be digitized, classified and made searchable. This is important for the protection and study of national linguistic and literary heritage.

Fifth, AI can support language education. It can be used to design exercises, analyze learner errors, provide feedback and create adaptive learning materials.

References:

1. UNESCO. (2023). *Guidance for Generative AI in Education and Research*. UNESCO.
2. Stanford Institute for Human-Centered Artificial Intelligence. (2025). *AI Index Report 2025*. Stanford University.
3. Fischer, T. (2024). Exploring Large Language Models for Qualitative Data Analysis Workflows. *Proceedings of NLP4DH 2024*. Association for Computational Linguistics.
4. Sullutrone, G. (2024). Large Language Models Integration in Digital Humanities. University of Modena and Reggio Emilia.
5. NLP4DH. (2024). *Natural Language Processing for Digital Humanities*.
6. Jurafsky, D., & Martin, J. H. (2024). *Speech and Language Processing*. Stanford University.
7. McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.
8. Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press.
9. Crystal, D. (2019). *The Cambridge Encyclopedia of the English Language*. Cambridge University Press.