BG Pulse Journal USA
ISSN:1947-2536 http://www.bgpulseusa.com/index.php/bg

THE ROLE OF ARTIFICIAL INTELLIGENCE AND LARGE LANGUAGE MODELS IN LINGUISTICS: THE CASE OF UZBEK AND ENGLISH LANGUAGES

Turakulova Nasiba Abduganiyevna
Teacher of the department of English language,
linguistics and literature
Samarkand State University named after Sharof Rashidov
turakulova908@gmail.com
https://doi.org/10.5281/zenodo.20372675

Abstract. Artificial intelligence has become an important part of modern linguistics and language-related technologies. In recent years, large language models have changed the way computers process and understand human language. These technologies are now widely used in machine translation, text generation, speech recognition, and language learning applications. This article explores the role of artificial intelligence and large language models in linguistics through a comparison of Uzbek and English languages. The study focuses on differences in linguistic structure, digital resources, and natural language processing performance. The findings show that English-language AI systems achieve higher accuracy because of extensive corpora and technological support, while Uzbek language technologies still face challenges related to limited datasets and morphological complexity. The article also discusses the future potential of artificial intelligence in the development of Uzbek computational linguistics and language technologies.

Keywords: Artificial intelligence, large language models, linguistics, natural language processing, computational linguistics, Uzbek language, English language, machine translation.

Introduction. In the twenty-first century, artificial intelligence (AI) has become one of the most rapidly developing technologies and has significantly influenced various scientific fields, including linguistics. The integration of AI into language studies has created new opportunities for analyzing, processing, and generating human language. In recent years, large language models (LLMs) have attracted particular attention because of their ability to produce human-like texts and perform complex natural language processing tasks (Brown et al., 2020). Large language models are based on deep learning techniques and transformer architecture, which allow computers to understand contextual relationships between words and sentences more effectively (Vaswani et al., 2017). These systems are now widely used in machine translation, speech recognition, text summarization, virtual assistants, and educational technologies. Applications such as ChatGPT (https://chatgpt.com/?utm_source=chatgpt.com) show how AI systems can generate fluent and meaningful responses that closely resemble human communication. English has become the leading language in artificial intelligence research due to the availability of extensive digital corpora, academic resources, and annotated datasets. Most AI systems are trained primarily on English-language materials, which increases their performance and accuracy in English natural language processing tasks (Jurafsky & Martin, 2023). Consequently, English-language AI systems demonstrate strong results in text generation, semantic analysis, and automated translation. In contrast, Uzbek language technologies are still developing in the field of computational linguistics.

One of the main reasons is the limited amount of linguistic data and the lack of large-scale corpora for training AI models. Moreover, Uzbek belongs to the agglutinative language family, where grammatical meanings are formed through multiple suffixes attached to root words. This

linguistic structure creates additional challenges for machine translation and text-processing systems. Despite these difficulties, interest in Uzbek-language AI technologies has increased in recent years. Researchers are working on improving digital corpora, creating NLP tools, and adapting large language models for Uzbek linguistic contexts. Therefore, examining the role of artificial intelligence and large language models in Uzbek and English linguistics is important for understanding both the current state and future development of computational linguistics.

Research methodology. This study employs a comparative and descriptive research design to investigate the role of artificial intelligence and large language models in Uzbek and English linguistics. The research is primarily based on qualitative analysis of natural language processing (NLP) tasks, including machine translation, text generation, semantic interpretation, and morphological processing. Comparative analysis was selected as the main method because it allows the identification of similarities and differences between AI performance in two linguistically distinct languages. The study examines outputs generated by large language models in both Uzbek and English. In addition, existing scientific literature related to computational linguistics, artificial intelligence, and natural language processing was reviewed to support the analysis (Jurafsky & Martin, 2023). AI systems based on transformer architecture were considered particularly relevant because transformers have become the foundation of modern language models (Vaswani et al., 2017). Several linguistic aspects were analyzed during the research process. First, the study compared the quality of machine translation between Uzbek and English. Particular attention was given to grammatical accuracy, contextual understanding, and sentence fluency. Previous studies suggest that AI-based translation systems perform more effectively in high-resource languages due to larger datasets and extensive training corpora (Koehn, 2020). Secondly, the research analyzed text generation capabilities in both languages. Outputs produced by large language models were examined to determine how naturally and coherently the systems generated texts in Uzbek and English. Since English dominates most AI training data, the study expected stronger performance in English-language text generation (Brown et al., 2020). Another important aspect of the methodology involved morphological analysis.

Uzbek is an agglutinative language in which grammatical meanings are expressed through suffixes, while English is primarily analytic and relies more heavily on word order. This structural difference was analyzed to determine how linguistic complexity influences AI performance in language processing tasks. Finally, descriptive analysis was used to interpret the collected data and evaluate the effectiveness of artificial intelligence technologies in both linguistic contexts. The methodology aimed to provide a balanced comparison of AI applications in Uzbek and English while identifying the major factors influencing computational language processing.

Results. The results of the study revealed noticeable differences in the performance of artificial intelligence systems in Uzbek and English language processing. These differences were mainly related to corpus availability, linguistic structure, and the amount of training data used in large language models.

Table 1. Comparison of AI Performance in Uzbek and English

Aspect	English	Uzbek
Machine translation accuracy	High	Moderate
Text generation fluency	Natural and coherent	Sometimes
Corpus availability	Extensive	Limited
Morphological complexity	Lower	Higher

NLP tool development	Highly developed	Developing

The analysis demonstrated that English-language processing achieved better overall performance in AI systems. English outputs were generally more fluent, contextually accurate, and grammatically stable. Uzbek-language outputs, while understandable, occasionally contained unnatural sentence structures and lexical repetition.

Machine Translation Analysis. Machine translation systems showed significantly higher performance in English due to the availability of large-scale training corpora and annotated datasets. Uzbek translations were generally accurate in simple sentences but became less natural in complex contexts.

Table 2. Sample Translation Comparison

Source Sentence	AI output
"Artificial intelligence improves language learning."	"Sun'iy intellekt til o'rganishni yaxshilaydi."
"Researchers are developing advanced language technologies."	"Tadqiqotchilar ilg'or til texnologiyalarini rivojlantirmoqda."

Although the translations above are grammatically correct, longer and more complex sentences occasionally produced structural inconsistencies in Uzbek outputs.

Example:

English source: "Despite limited digital resources, Uzbek computational linguistics continues to develop rapidly."

AI-generated Uzbek output: "Cheklangan raqamli resurslarga qaramay, o'zbek kompyuter lingvistikasi tez rivojlanishda davom etmoqda."

The translation preserves the general meaning, but the sentence structure is slightly less natural compared to English-generated outputs.

Text Generation Performance. Large language models demonstrated stronger text generation capabilities in English than in Uzbek. English texts were more coherent, stylistically varied, and contextually detailed.

Table 3. Text Generation Characteristics

Feature	**English output**	**Uzbek output**
Sentence fluency	High	Moderate
Vocabulary diversity	Wide	Limited
Context preservation	Strong	Partial
Repetition frequency	Low	Moderate

For example, AI-generated English paragraphs usually contained more natural transitions between ideas, whereas Uzbek texts sometimes relied on repetitive sentence patterns.

Morphological Processing. Another important finding involved morphological analysis. Uzbek language structure created additional processing difficulties because grammatical meanings are expressed through multiple suffixes attached to root words.

Table 4. Morphological Complexity

Language	**Linguistic type**	**AI processing difficulty**
English	Analytic	Lower
Uzbek	Agglutinative	Higher

For instance, the Uzbek word:

“o‘qituvchilarimiznikidan”

contains several grammatical suffixes within a single lexical unit. Such structures may increase ambiguity for AI systems during tokenization and semantic interpretation. In contrast, English usually expresses grammatical relationships through separate words and fixed word order, which simplifies computational processing.

Overall, the findings indicate that large language models currently perform more effectively in English than in Uzbek. Nevertheless, the study also demonstrates that Uzbek-language AI technologies are gradually improving as more digital resources and linguistic datasets become available.

Discussion. The findings of this study demonstrate that artificial intelligence and large language models have a significant impact on modern linguistics, particularly in the field of natural language processing. However, the effectiveness of AI systems differs considerably between English and Uzbek due to linguistic, technological, and resource-related factors. One of the main reasons for the higher performance of English-language AI systems is the availability of extensive digital corpora and annotated datasets. Most modern language models are trained primarily on English-language materials, which improves their ability to generate fluent and contextually accurate outputs (Brown et al., 2020). As a result, AI systems demonstrate stronger performance in English-language tasks such as machine translation, semantic analysis, and automated text generation. In contrast, Uzbek remains a low-resource language in computational linguistics. The lack of large-scale corpora and standardized datasets directly affects the quality of AI-generated outputs. During the analysis, Uzbek-language responses occasionally showed lexical repetition, structural simplification, and contextual limitations. Similar challenges for low-resource languages have also been discussed in previous NLP studies (Jurafsky & Martin, 2023). Another important factor influencing AI performance is linguistic structure.

English belongs to the analytic language group and relies heavily on word order to express grammatical relationships. Uzbek, however, is an agglutinative language in which grammatical meanings are formed through multiple suffixes attached to root words. This increases morphological complexity and creates additional challenges for machine learning systems during tokenization and semantic interpretation (Manning & Schütze, 1999). The study also revealed that large language models are more effective when processing languages with richer training environments. Transformer-based architectures introduced by Vaswani et al. (2017) significantly improved contextual understanding in NLP systems, but their efficiency still depends largely on the quantity and quality of available training data.

Consequently, English-language outputs are usually more natural and stylistically diverse than Uzbek-generated texts. Despite these limitations, the findings indicate that Uzbek computational linguistics has strong development potential. In recent years, increasing attention has been given to creating Uzbek digital corpora, improving translation systems, and adapting AI technologies to local linguistic structures. Expanding language datasets and developing specialized NLP tools could substantially improve the quality of Uzbek-language AI systems in the future.

Conclusion. This study examined the role of artificial intelligence and large language models in linguistics through a comparative analysis of Uzbek and English languages. The results demonstrated that AI systems currently perform more effectively in English because of extensive linguistic corpora, advanced computational resources, and broader research support. At the same time, Uzbek-language processing still faces several challenges related to limited datasets, insufficient annotated corpora, and morphological complexity. These factors affect the fluency,

contextual accuracy, and overall effectiveness of AI-generated outputs in Uzbek. Nevertheless, the study suggests that the future of Uzbek computational linguistics is promising. The continuous development of digital resources, NLP technologies, and language-specific AI models may significantly improve Uzbek-language processing in the coming years. In a nutshell, artificial intelligence and large language models are transforming modern linguistics by creating new opportunities for language analysis, communication, and education. Their future effectiveness, however, will strongly depend on the development of linguistic resources and technological infrastructure for both global and low-resource languages.

References:

1. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Proceedings of the ACM Conference on Fairness, Accountability, and Transparency.
2. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media.
3. Brown, T. B., et al. (2020). Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems (NeurIPS).
4. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT.
5. Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. Minds and Machines, 30(4), 681–694.
6. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
7. Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processing (3rd ed.). Stanford University.
8. Koehn, P. (2020). Neural Machine Translation. Cambridge University Press.
9. Manning, C. D., & Schütze, H. (1999). Foundations of Statistical Natural Language Processing. MIT Press.
10. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781.
11. OpenAI. (2023). GPT-4 Technical Report. arXiv preprint arXiv:2303.08774.
12. Vaswani, A., et al. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems (NeurIPS).