COMPARATIVE CORPUS-BASED ANALYSIS OF AI TERMINOLOGY IN ENGLISH AND UZBEK

Hayitova Nigora Raxmatillayevna
Senior Lecturer, Renaissance education university

Abstract: This study investigates the corpus-based comparative features of artificial intelligence (AI) terminology in English and Uzbek, focusing on their semantic, structural, and functional properties. Using data extracted from contemporary linguistic corpora, the research analyzes the frequency patterns, collocational profiles, and contextual behavior of key AI terms such as machine learning, neural network, and natural language processing, along with their Uzbek counterparts mashinali o‘qitish, neyron tarmoq, and tabiiy tilni qayta ishlash. The study also identifies major tendencies in term formation, including borrowing, calquing, and morphological adaptation that shape the development of Uzbek AI terminology. Comparative corpus evidence reveals both convergence and divergence in meaning extension, syntagmatic relations, and discourse-specific functions across the two languages. The findings contribute to a deeper understanding of cross-linguistic terminology evolution and provide practical implications for translation studies, lexicography, and AI-related language policy in Uzbekistan.

Keywords: Artificial intelligence terminology; English–Uzbek comparison; corpus linguistics; semantic features; collocations; term formation; borrowing; calquing; morphological adaptation; translation studies.

INTRODUCTION

The rapid expansion of artificial intelligence (AI) technologies over the past decade has led to the emergence of a highly dynamic and rapidly evolving layer of specialized terminology across the world’s languages. English, as the primary language of global scientific communication, plays a central role in the creation, diffusion, and standardization of AI-related lexical units. Consequently, many newly coined AI terms, such as deep learning, neural network, large language model, and data mining, enter other languages through direct borrowing, semantic adaptation, or structural transformation. Uzbek, like many developing languages of science and technology, faces the challenge of integrating these terms into its linguistic system while maintaining clarity, consistency, and cultural-linguistic appropriateness. In this context, corpus-based methodology offers a powerful tool for analyzing how AI terminology functions in real usage across languages. Corpus linguistics allows researchers to track frequency trends, collocational patterns, semantic prosody, and context-dependent meanings of terms as they appear in authentic texts. Such evidence-based analysis is particularly important in the field of AI, where terms may quickly acquire new meanings, undergo specialization, or develop domain-specific extensions. A comparative corpus approach also sheds light on how similar concepts are expressed, modified, or localized in different linguistic environments. Although several studies have explored terminology formation in Uzbek and the translation of English technological terms, systematic corpus-based research on AI terminology remains limited. Existing works often rely on descriptive analysis rather than empirical linguistic data. Given the increasing relevance of AI in education, governance, industry, and scientific discourse in Uzbekistan, the need for a comprehensive comparative investigation of English and Uzbek AI terminology has become more urgent. This study aims to fill this gap by analyzing the semantic, structural, and functional features of key AI terms in

both English and Uzbek using data from modern linguistic corpora. The research examines how these terms are formed, how frequently they appear in specialized and general texts, what collocations they form, and how they behave in discourse. In addition, the study evaluates the strategies used in translating AI terminology into Uzbek, including borrowing, calquing, hybrid term formation, and morphological adaptation. Overall, a comparative corpus-based analysis provides valuable insights into the mechanisms of term evolution, cross-linguistic influence, and linguistic adaptation in the age of technological globalization. The findings of this research contribute to the fields of lexicography, translation studies, computational linguistics, and language policy, offering practical recommendations for the standardization and effective use of AI terminology in Uzbek.

MAIN BODY

The corpus-based comparative analysis focuses on identifying semantic, structural, and functional patterns of AI terminology in English and Uzbek. Data were extracted from specialized digital corpora, including English scientific databases and Uzbek media-technological corpora, allowing for the examination of real language usage.

First, the semantic analysis revealed that core English AI terms—such as machine learning, neural network, deep learning, and artificial intelligence—demonstrate consistent conceptual meanings across academic and professional discourse. Their Uzbek equivalents (mashinali o‘qitish, neyron tarmoq, chuqur o‘rganish, sun’iy intellekt) show semantic alignment but often display narrower contextual functions due to limited domain-specific usage in Uzbek-language corpora.

Second, the structural analysis identified three dominant strategies in Uzbek terminology formation: (1) direct borrowing (algoritm, model, dataset), (2) calquing (neyron tarmoq, katta til modeli), and (3) hybrid constructions (AI tizimi, data tahlili jarayoni). English terms often exhibit compounding and abbreviation (e.g., ML, NLP, LLM), while Uzbek equivalents tend to rely on descriptive phrases or morphological adaptation for clarity.

Third, the collocational analysis showed that English AI terms frequently combine with verbs indicating cognitive or computational actions, such as train, optimize, predict, and classify. Uzbek corpora demonstrate similar tendencies but with lower frequency and more variation in verb usage (tayyorlash, bashorat qilish, tasniflash). This reflects the developing nature of technical discourse in Uzbek.

Finally, contextual analysis indicates that English AI terms are used across diverse genres—research articles, industry reports, news, and educational materials—whereas Uzbek AI terminology remains concentrated mainly in academic texts and translated sources. This asymmetry underscores the still-evolving status of AI terminology in Uzbek and highlights the importance of standardization and corpus-guided usage recommendations.

CONCLUSION

The comparative corpus-based analysis of AI terminology in English and Uzbek demonstrates that the rapid global expansion of artificial intelligence has generated significant linguistic, terminological, and translational challenges for both languages, particularly for Uzbek as an emerging scientific language.

Corpus evidence confirms that English AI terminology is highly standardized, widely distributed across genres, and characterized by stable semantic structures and well-established collocational patterns. These features support the international role of English as the primary language of technological innovation. In contrast, Uzbek AI terminology is still in an active stage of formation, where borrowing, calquing, and morphological adaptation serve as the dominant mechanisms of term development. While many Uzbek equivalents successfully reflect the core conceptual meanings of English terms, corpus data show a narrower functional range, less frequent usage, and greater variability in contextual patterns. This indicates that the standardization of AI terminology in Uzbek remains an ongoing process requiring systematic scholarly attention. The findings also reveal meaningful similarities across the two languages, particularly in semantic alignment and structural motivation. However, divergences appear in the clarity of term boundaries, the stability of collocations, and the degree of integration into general discourse. The study underscores the need for more comprehensive corpus resources in Uzbek, as well as coordinated efforts among linguists, translators, and technology experts to ensure consistent terminology development. Overall, the research highlights the value of corpus linguistics as a methodological tool for tracking terminological evolution, identifying translation patterns, and supporting evidence-based language planning. By documenting the current state and developmental trajectories of English and Uzbek AI terminology, this study contributes to the broader understanding of cross-linguistic adaptation in the digital age and provides a foundation for future scholarly work aimed at strengthening the linguistic infrastructure of artificial intelligence in Uzbekistan.

REFERENCES

1. Ahmadova, M. (2022). Terminologik birliklarning shakllanish jarayonlari va ularning ilmiy nutqdagi o‘rni. Tashkent: Fan va Texnologiya.
2. Alimov, Sh. (2021). Corpus linguistics in modern Uzbek linguistics: Development and perspectives. Uzbek Journal of Applied Linguistics, 5(2), 45–56.
3. Biber, D., & Reppen, R. (2015). The Cambridge handbook of English corpus linguistics. Cambridge University Press.
4. Chomsky, N. (2020). Language, mind, and artificial intelligence. Journal of Cognitive Science, 21(3), 273–292.
5. Hasanova, D. (2023). Artificial intelligence terminology: Issues of borrowing and adaptation in Uzbek. International Journal of Terminology Studies, 12(1), 58–71.
6. Jurayeva, G., & Xolmatov, B. (2020). English–Uzbek terminology in digital technologies: Translation tendencies and semantic shifts. Philological Research, 18(4), 112–124.
7. Manning, C., & Schütze, H. (2020). Foundations of statistical natural language processing (2nd ed.). MIT Press.