STRUCTURAL-SEMANTIC AND COMPARATIVE ANALYSIS OF ARTIFICIAL INTELLIGENCE TERMS IN THE UZBEK AND ENGLISH LANGUAGES

Karimova Nurkhon Tulkinovna
Independent researcher at Navoi State University
doi:https://doi.org/10.5281/zenodo.22831672

Annotation: This article presents a structural-semantic and comparative analysis of artificial intelligence (AI) terminology in the Uzbek and English languages. Using a corpus of 150 AI-related terms, the study classifies terms by structural model (simple, derived, compound, phrasal) and semantic field (algorithms, data, machine learning, neural networks, robotics), revealing similarities and differences rooted in the typological contrast between agglutinative Uzbek and analytic-synthetic English. The findings show that English AI terminology is dominated by compounding and blending, while Uzbek relies mainly on borrowing, calquing, and affixation. Recommendations for standardizing Uzbek AI terminology are offered.

Keywords: artificial intelligence, terminology, structural-semantic analysis, comparative linguistics, Uzbek language, English language, neologism, calque, borrowing, word-formation.

The rapid advancement of artificial intelligence (AI) technologies over the past two decades has generated an extensive and constantly evolving terminological system. Concepts such as machine learning, deep learning, neural network, natural language processing, and big data have entered not only specialized scientific discourse but also everyday language use. This terminological expansion presents a unique challenge for languages with different typological structures, particularly for languages such as Uzbek, which must adapt or create equivalents for concepts originally formulated in English.

The English language, functioning as the primary medium of international scientific and technological communication, generates the majority of new AI terms. Uzbek, in turn, faces the task of either borrowing these terms directly, transliterating them, or creating native equivalents through translation and word-formation processes. This dynamic makes the comparative study of AI terminology in Uzbek and English particularly relevant, both for theoretical linguistics and for practical fields such as translation studies, terminography, and language planning.

The relevance of this research is determined by several factors. First, the absence of a unified, standardized system of AI terminology in the Uzbek language creates inconsistency in scientific literature, educational materials, and mass media. Second, the structural-semantic peculiarities of term formation in agglutinative languages such as Uzbek have not been sufficiently studied in relation to rapidly emerging technological fields. Third, a systematic comparison with English can help identify effective strategies for term creation and adaptation that preserve semantic precision while conforming to the grammatical norms of the Uzbek language.

The aim of this article is to conduct a structural-semantic and comparative analysis of AI terminology in Uzbek and English, to identify the dominant word-formation models in each language, and to classify the semantic relations that characterize this specialized lexical layer. The objectives of the study include: (1) reviewing theoretical approaches to terminology and term formation; (2) describing the methodology and corpus used for analysis; (3) analyzing the structural types of AI terms in both languages; (4) examining semantic relations such as polysemy, synonymy, and hyponymy within the AI terminological field; and (5) comparing the two systems to reveal typological similarities and differences.

Terminology as a linguistic discipline has been extensively studied by scholars such as Eugen Wüster, who laid the theoretical foundations of terminology science, and later researchers who expanded this framework to cover domain-specific lexicons in rapidly developing fields. Structural approaches to term formation typically distinguish between simple terms (single-root words), derived terms (formed through affixation), compound terms (formed by combining two or more roots), and phrasal or multi-word terms (terminological word combinations).

In English linguistics, the study of scientific and technical terminology has emphasized compounding, blending, and acronymy as the primary means of term formation, particularly in computer science and information technology. Terms such as chatbot (chat + robot), deepfake (deep + fake), and machine learning (a compound noun phrase) illustrate the productivity of these processes.

In Uzbek linguistics, terminology studies have traditionally focused on borrowings from Russian and, more recently, from English, alongside native word-formation strategies based on the agglutinative structure of the language. Scholars of Uzbek terminology have noted the tendency toward calquing (loan translation) and semantic extension of existing native lexemes to express new technical concepts, alongside direct phonetic borrowing of internationalisms. However, the specific field of AI terminology remains underexplored in Uzbek linguistic scholarship, which justifies the need for a dedicated structural-semantic investigation such as the present study.

Comparative and contrastive terminology studies, which examine how equivalent concepts are lexicalized across typologically different languages, provide the methodological foundation for this article. Such studies are particularly valuable for translation practice, terminology standardization, and the development of bilingual technical dictionaries.

The research is based on a corpus of 150 AI-related terms selected from English-language scientific publications (IEEE, ACM, and Springer AI journals), technical documentation, and terminological glossaries, together with their Uzbek equivalents extracted from Uzbek-language scientific articles, university course materials, and official terminological resources published by the State Language Committee and relevant ministries.

The following methods were applied: (1) the descriptive method, for characterizing the structural and semantic properties of the terms; (2) the comparative-contrastive method, for identifying similarities and differences between the Uzbek and English terminological systems; (3) componential (semantic) analysis, for decomposing term meanings into semantic components; and (4) quantitative analysis, for calculating the frequency distribution of structural types within each language. Terms were classified into four structural categories — simple, derivative, compound, and phrasal/multi-word — and into five semantic fields: algorithms and computation, data and data processing, machine learning and training, neural networks and architectures, and robotics and automation.

The structural analysis of the corpus reveals distinct patterns of term formation in English and Uzbek, largely conditioned by the typological characteristics of the two languages. English, as an analytic-synthetic language with flexible compounding rules, favors multi-word and compound terms, while Uzbek, as an agglutinative language, relies more heavily on affixation and borrowing.

Analysis of the English corpus shows that compound and phrasal terms constitute the largest structural category, accounting for the majority of AI-related vocabulary. Examples include neural network, machine learning, natural language processing, and reinforcement learning, all of which combine two or more independent lexical units into a single terminological unit. A second productive category is blending, exemplified by terms such as chatbot and deepfake, where two words are fused into one. Simple and derived terms, such as algorithm, classifier, and tokenization, are comparatively less frequent but remain essential to the core vocabulary of the field.

4.2. Structural Types in Uzbek AI Terminology

In the Uzbek corpus, direct borrowings and phonetically adapted internationalisms dominate, such as algoritm (algorithm), robot (robot), and neyron tarmoq (neural network, a calque). Derivational models using native Uzbek suffixes are also productive, for example, o‘qitish (training, from the verb o‘qitmoq with a deverbal noun suffix) and tasniflovchi (classifier, formed with the agentive suffix -uvchi/-lovchi). Compound and calque formations, such as sun’iy intellekt (artificial intelligence) and chuqur o‘rganish (deep learning), represent an intermediate strategy in which the semantic structure of the English source term is translated using native morphological resources rather than being borrowed phonetically.

The comparative distribution of structural types across the two languages is summarized in Table 1.

Structural type	English (%)	Uzbek (%)
Simple terms	12%	18%
Derived terms (affixation)	18%	27%
Compound terms	34%	15%
Phrasal / multi-word terms	28%	12%
Direct borrowings / calques	8%	28%

As Table 1 demonstrates, English shows a clear predominance of compound and phrasal structures (62% combined), reflecting the language's syntactic flexibility in forming multi-element technical terms. Uzbek, by contrast, shows a higher proportion of derived terms and borrowings/calques (55% combined), reflecting both its agglutinative morphology, which favors suffixation over compounding, and its position as a receiving language for internationally established AI concepts.

The semantic analysis of the corpus focuses on the organization of AI terminology into conceptual fields and on the semantic relations — synonymy, polysemy, and hyponymy — that structure this specialized vocabulary in both languages.

Both English and Uzbek AI terminology can be grouped into five major semantic fields identified in the corpus: (1) algorithms and computation (e.g., algorithm/algoritm, heuristic/evristika); (2) data and data processing (e.g., dataset/ma'lumotlar to‘plami, big data/katta ma'lumotlar); (3) machine learning and training (e.g., supervised learning/nazoratli o‘qitish, overfitting/haddan tashqari moslashuv); (4) neural networks and architectures (e.g., neural network/neyron tarmoq, convolutional layer/konvolyutsion qatlam); and (5) robotics and automation (e.g., robot/robot, autonomous system/avtonom tizim). This shared conceptual structure confirms that the underlying scientific ontology of AI is largely universal, while its lexical realization differs according to language-specific word-formation strategies.

Synonymy is observed in both languages, particularly during the early stage of term adoption when multiple equivalents compete for a single concept. For instance, the Uzbek equivalent of artificial intelligence appears as both sun’iy intellekt (calque) and sometimes as the borrowed sun’iy aql, before terminological standardization narrows usage to a single preferred form. In English, similar

competition exists between terms such as machine intelligence and artificial intelligence in early literature.

Polysemy is less common in core AI terms, as terminological systems generally favor monosemy for precision; however, certain general-language words acquire specialized AI meanings, creating a form of terminological polysemy. The English word model, for example, carries a specialized meaning in machine learning (a trained computational representation) distinct from its everyday sense. The corresponding Uzbek term model exhibits the same semantic extension, indicating a parallel process of terminologization through calquing of the English semantic structure rather than independent semantic development.

Hyponymic relations are prominent in both terminological systems, reflecting the hierarchical structure of AI as a scientific domain. In English, artificial intelligence functions as a hyperonym for hyponyms such as machine learning, which itself is a hyperonym for deep learning, reinforcement learning, and supervised learning. The Uzbek system mirrors this hierarchy precisely — sun'iy intellekt > mashinali o'rganish > chuqur o'rganish — demonstrating that hyponymic relations are typically preserved through calquing rather than restructured according to native conceptual categories.

The comparative analysis reveals both convergence and divergence between the two terminological systems. Convergence is evident at the semantic level: the conceptual organization of AI terminology into fields such as algorithms, data, learning, and neural architectures is essentially identical in Uzbek and English, since Uzbek terminology largely develops by translating or borrowing the internationally established English-based conceptual system rather than generating an independent ontology.

Divergence, however, is pronounced at the structural level. English favors compounding and phrase formation, exploiting its relatively free syntactic combinability to create new multi-word terms with minimal morphological marking. Uzbek, constrained by its agglutinative structure, cannot easily replicate English-style noun compounding and instead relies on a combination of (a) direct phonetic borrowing for internationally standardized concepts (algoritm, robot, klaster), (b) calquing for transparent compound concepts (sun'iy intellekt, chuqur o'rganish), and (c) native affixation for process- and agent-denoting terms (o'qitish, tasniflovchi, aniqlovchi).

This distribution suggests that Uzbek terminology formation in the AI domain follows a mixed strategy that balances international intelligibility (through borrowing) with linguistic accessibility (through calquing and native derivation). The relatively high proportion of unassimilated borrowings, however, also points to a gap in systematic terminological planning, since inconsistent transliteration practices (for example, variation between neyron tarmoq and neyral tarmoq) can create confusion in scientific and educational contexts.

The structural-semantic and comparative analysis of artificial intelligence terminology in Uzbek and English demonstrates that while the two languages share a largely common conceptual and semantic organization of the AI domain, they diverge significantly in the structural mechanisms used to encode these concepts lexically. English relies predominantly on compounding, blending, and phrasal term formation, whereas Uzbek combines direct borrowing, calquing, and native affixation, reflecting its agglutinative typology and its role as a receiving language in this technological field.

These findings have practical implications for terminology planning, translation, and lexicography in Uzbekistan. Standardizing the transliteration and calquing practices used for AI terms would reduce variation and improve the consistency of scientific and educational discourse. Furthermore, the systematic use of native Uzbek derivational morphology, where semantically transparent and pedagogically effective, can enhance the accessibility of AI education for Uzbek-speaking learners while maintaining compatibility with international terminological standards. Future research could expand the corpus to include emerging subfields such as generative AI and large language models, and

could incorporate corpus-based frequency analysis of terminological usage in Uzbek digital media and academic publishing.

References

1. Cabré, M. T. (1999). Terminology: Theory, Methods and Applications. John Benjamins Publishing.
2. Sager, J. C. (1990). A Practical Course in Terminology Processing. John Benjamins Publishing.
3. Bekmurodov, A. (2018). O‘zbek tili terminologiyasi masalalari. Toshkent: Fan.
4. Mengliyev, B. (2017). O‘zbek tilida lug‘at boyitish jarayonlari. Toshkent: O‘qituvchi.
5. Russell, S., & Norvig, P. (2020). Artificial Intelligence: A Modern Approach (4th ed.). Pearson.
6. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
7. Yusupov, F. (2021). Ilmiy-texnikaviy terminologiyaning shakllanish xususiyatlari. O‘zbek tili va adabiyoti jurnali, 4, 45–52.