SCIENCE CONNECTS PEOPLE

INTERNATIONAL SCIENTIFIC JOURNAL OF

2026 / No. 1(1)

Contemporary Pedagogical and Philological Research

UDK: 81'322:81'42:004.8

LINGUISTIC MARKERS OF AI-GENERATED TEXTS IN UZBEK: A STYLOMETRIC AND PRAGMATIC ANALYSIS OF HUMAN-WRITTEN AND LLM-GENERATED ACADEMIC DISCOURSE.

Muradov Mukhammad Obomislim ogli

Teacher, Department of Interfaculty Foreign Languages, Faculty of World Languages, Namangan State University. ORCID: 0009-0002-1461-1833.

Annotation. This article examines the linguistic markers of LLM-generated Uzbek academic texts from stylometric and pragmatic perspectives. The relevance of the study is determined by the rapid spread of AI-based text generation in scholarly communication, the growing importance of academic integrity and authorship attribution, and the lack of research on this issue in Uzbek as a low-resource language. The aim of the study is to identify and describe the lexica-grammatical, syntactic and pragmatic features that distinguish human-written academic discourse from texts generated by large language models. The material comprises a comparative corpus of human-authored and LLM-generated texts on identical topics. The analysis employs stylometric methods (lexical diversity, mean sentence and word length, word frequency distribution), discourse-pragmatic analysis (hedging, metadiscourse markers, modality, expression of authorial stance) and statistical testing. The findings reveal distinctive linguistic “fingerprints” of LLM-generated texts, including stylistic uniformity, lexical repetition, calqued syntactic structures and pragmatic neutrality. The scientific novelty lies in the systematic description of linguistic markers of AI-generated texts in Uzbek. The results can be applied to authorship attribution, academic integrity policies, the development of automatic detection tools for Uzbek, and philological education.

Keywords: artificial intelligence, large language models (LLMs), linguistic markers, stylometry, pragmatics, academic discourse, corpus analysis, hedging, meta discourse, Uzbek language.

UDK: 81'322:81'42:004.8

O‘ZBEK TILIDA SUN’IY INTELLEKT YARATGAN MATNLARNING LINGVISTIK MARKERLARI AKADEMIK DISKURSNING STILOMETRIK VA PRAGMATIK TAHLILI.

Muradov Muxammad Obomislim o‘g‘li

Namangan davlat universiteti, Jahon tillar fakulteti, Fakultetlararo chet tillar tillar kafedrasi, o‘qituvchisi. ORCID: 0009-0002-1461-1833.

Annotatsiya. Maqolada katta til modellari (LLM) tomonidan yaratilgan o‘zbekcha akademik matnlarning lingvistik markerlari stilometrik va pragmatik yondashuv asosida tahlil qilinadi. Tadqiqotning dolzarbligi sun’iy intellekt asosidagi matn generatsiyasining ilmiy muloqotga keng kirib borishi, akademik halollik va muallifni aniqlash masalalarining keskinlashuvi, shuningdek, o‘zbek tili kam resursli til sifatida bu yo‘nalishda yetarlicha o‘rganilmaganligi bilan belgilanadi. Maqolaning maqsadi inson yozgan va LLM yaratgan akademik matnlarni farqlovchi leksik-grammatik, sintaktik va pragmatik ko‘rsatkichlarni aniqlash va tavsiflashdan iborat. Tadqiqot materiali sifatida bir xil mavzudagi inson muallif matnlari va LLM generatsiya qilgan matnlardan tuzilgan qiyosiy korpus olindi. Tahlilda stilometrik usullar (leksik xilma-xillik, o‘rtacha gap va so‘z uzunligi, so‘z chastotasi taqsimoti), diskurs va pragmatik tahlil (hedging, metadiskurs markerlari, modallik, muallif pozitsiyasining ifodalanishi) hamda statistik tekshiruv usullari qo‘llanildi. Natijalar LLM matnlarining o‘ziga xos lingvistik “izlari”ni, jumladan, uslubiy bir xillik, leksik takrorlar, kalkalangan sintaktik qurilmalar va pragmatik betaraflikni aniqlash imkonini beradi. Ilmiy yangilik shundaki, o‘zbek tilida sun’iy intellekt matnlarining lingvistik markerlari tizimli tavsiflanadi. Tadqiqot natijalaridan matn muallifligini aniqlash, akademik halollikni ta’minlash, o‘zbek tili uchun avtomatik aniqlash vositalarini yaratish va filologik ta’lim jarayonida foydalanish mumkin.

Kalit so‘zlar: sun’iy intellekt, katta til modellari (LLM), lingvistik markerlar, stilometriya, pragmatika, akademik diskurs, korpus tahlili, hedging, metadiskurs, o‘zbek tili.

УДК: 81'322:81'42:004.8

ЛИНГВИСТИЧЕСКИЕ МАРКЕРЫ ТЕКСТОВ, СГЕНЕРИРОВАННЫХ ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ, В УЗБЕКСКОМ ЯЗЫКЕ: СТИЛОМЕТРИЧЕСКИЙ И ПРАГМАТИЧЕСКИЙ АНАЛИЗ АКАДЕМИЧЕСКОГО ДИСКУРСА.

Мурадов Мухаммад Обомислимович

Преподаватель английского языка, кафедры межфакультетских иностранных языков, факультета мировых языков Наманганского государственного университета. ORCID: 0009-0002-1461-1833.

Аннотация. В статье анализируются лингвистические маркеры академических текстов на узбекском языке, сгенерированных большими языковыми моделями (LLM), с позиций стило метрического и прагматического подходов. Актуальность исследования обусловлена широким проникновением генерации текстов на основе искусственного интеллекта в научную коммуникацию, обострением проблем академической честности и атрибуции авторства, а также недостаточной изученностью данной проблематики на материале узбекского языка как мало ресурсного. Цель работы заключается в выявлении и описании лексико-грамматических, синтаксических и прагматических признаков, отличающих академический дискурс, созданный человеком, от текстов, сгенерированных LLM. Материалом послужил сопоставительный корпус текстов, написанных человеком и сгенерированных LLM на одинаковые темы. В работе использованы стило метрические методы (лексическое разнообразие, средняя длина предложения и слова, распределение частотности лексики), дискурсивно-прагматический анализ (хеджирование, мета

SCIENCE CONNECTS PEOPLE

дискурсивные маркеры, модальность, выражение авторской позиции) и статистическая проверка. Результаты позволяют выявить специфические лингвистические «следы» LLM-текстов: стилистическую однородность, лексические повторы, калькированные синтаксические конструкции и прагматическую нейтральность. Научная новизна состоит в системном описании лингвистических маркеров текстов искусственного интеллекта в узбекском языке. Результаты могут быть использованы при установлении авторства, обеспечении академической честности, разработке инструментов автоматического выявления для узбекского языка и в филологическом образовании.

Ключевые слова: искусственный интеллект, большие языковые модели (LLM), лингвистические маркеры, стило метрия, прагматика, академический дискурс, корпусный анализ, хеджирование, мета дискурс, узбекский язык.

Introduction.

Since the public release of ChatGPT in November 2022, large language models (LLMs) have become a routine instrument of scholarly writing. An analysis of more than 15 million PubMed abstracts from 2010–2024, based on shifts in “excess vocabulary”, suggests that at least 13.5% of 2024 abstracts were processed with LLMs, and that the effect on scientific prose has been unprecedented, exceeding that of the COVID-19 pandemic [1, art. eadt3813]. Academic discourse, traditionally regarded as the most regulated and conventionalized register of written language, is therefore being reshaped by machine-generated text, which raises both linguistic and ethical questions.

The relevance of the present study is determined by three interrelated problems. First, academic integrity: universities and journals increasingly need reliable ways to identify machine-generated text. Second, authorship attribution: automatic detectors are not neutral instruments. Seven widely used GPT detectors misclassified, on average, 61.3% of essays written by non-native English speakers as AI-generated, because they

rely on text perplexity, which penalizes writers with a restricted range of vocabulary and constructions [3, art. 100779]. Detection without a linguistic description of what distinguishes human and machine texts is therefore both unreliable and potentially discriminatory. Third, the low-resource status of many languages: the distribution of language resources in NLP is extremely skewed, and most languages remain marginal in language technology research [4, pp. 6282–6293]. Uzbek, an agglutinative Turkic language, belongs to this group. Surveys of Central Asian Turkic languages point to data scarcity and limited linguistic resources [5, pp. 309–324], and the available Uzbek NLP work is largely devoted to building pretrained models and evaluating them on tasks such as sentiment analysis, topic classification and named entity recognition [6, pp. 33–44].

The aim of the article is to identify and describe the linguistic markers that differentiate human-written and LLM-generated Uzbek academic texts. The aim is achieved through the following tasks: (1) to compile a comparative corpus of human-written and LLM-generated academic texts on identical topics; (2) to determine stylometric indicators of both sub corpora (lexical diversity, sentence and word length, frequency distribution); (3) to analyses pragmatic markers of stance and engagement (hedging, boosting, self-mention, attitude markers); (4) to build a typology of characteristic errors and stylistic “traces” of machine-generated texts, such as calques, unnatural word order and semantic vagueness.

The scientific novelty lies in the first systematic stylometric and pragmatic description of LLM-generated academic discourse in Uzbek and in the proposed typology of its linguistic markers. The practical significance is that the results can be used in authorship verification and academic integrity policies, in developing detection tools that take Uzbek morphology into account, in teaching academic writing, and in compiling corpus-based materials for Uzbek philology.

SCIENCE CONNECTS PEOPLE

The article is structured as follows: Section 2 reviews the literature on stylometry, LLM text analysis and academic pragmatics; Section 3 describes the corpus and methods; Section 4 presents the results; Section 5 discusses them; Section 6 concludes.

Literature Review.

Modern quantitative authorship attribution is conventionally traced to Mosteller and Wallace's study of the disputed Federalist Papers, which replaced expert judgement with statistical inference based on word frequencies [7; 9, p. 538]. For the following decades the field was dominated by attempts to define features that quantify writing style, and only later turned to objective evaluation and to short, noisy texts [9, pp. 538–556]. A landmark in this tradition is Burrows' Delta, which compares texts through the standardized relative frequencies of the most common words: the candidate whose texts show the smallest mean difference from the test text has the strongest claim to authorship [8, p. 271]. Burrows himself pointed out that the procedure is dependable for texts exceedingly roughly 1,500 words [8, p. 267]. Later work has clarified why Delta works. Evert et al. demonstrated that normalizing feature vectors to uniform length, which is implicit in the cosine measure, is the decisive factor behind its recent improvements [11, pp. ii4–ii16]. Two points matter for the present study. First, these methods were designed to discriminate between candidate human authors, so their application to the human versus LLM contrast requires a different interpretation of the "author" category. Second, the text-length requirement is a practical constraint for corpora of short academic texts such as abstracts, and it has to be accounted for in the methodology.

Academic writing is no longer regarded as impersonal and purely informational, but as a form of interaction between writer and reader. Hyland's work provides the principal descriptive apparatus: hedging as the expression of epistemic caution in research articles [15], meta discourse as the linguistic means of organizing interaction [16], and the model of stance and engagement [10, p. 173]. In this model, stance covers

hedges, boosters, attitude markers and self-mention, while engagement concerns how writers relate to readers with respect to the positions advanced in the text [10, p. 176]. Register-based approaches complement this view. Biber's multidimensional analysis identifies dimensions of variation through co-occurring linguistic features [17], and the register perspective links such features to situational characteristics [18]. Recent studies have applied these frameworks to LLM output. A comparison of 145 ChatGPT essays with 145 student essays found that students used significantly more engagement markers and a greater variety of them, particularly questions and personal asides, whereas ChatGPT relied more on appeals to shared knowledge [14, pp. 463–492]. Together with the finding on epistemic markers in [12], this suggests that LLM-generated academic prose tends to be interactionally flatter. Yet this evidence comes from English student writing, and marker inventories cannot be transferred directly to a language that realises epistemic stance through its own particles, modal words and verbal morphology.

The review reveals four conclusions. First, stylometric methods are well developed but were built for attributing human authorship and depend on text length [8; 11]. Second, comparative studies of human and LLM text are overwhelmingly based on English, and their findings on lexical diversity are inconsistent [12; 13]. Third, the pragmatic evidence points to reduced epistemic and engagement marking in LLM output [12; 14], but it concerns English student essays rather than academic discourse in a typologically different language. Fourth, Uzbek research has concentrated on resources and models rather than on the linguistic description of AI-generated text [5; 6; 19; 20]. The present study addresses this gap by combining stylometric and pragmatic analysis on a comparative corpus of human-written and LLM-generated Uzbek academic texts.

Methodology.

SCIENCE CONNECTS PEOPLE

The study follows a comparative, corpus-based, mixed-methods design. The quantitative strand extracts lexical, syntactic, morphological and pragmatic indicators from two sub corpora (human-written and LLM-generated Uzbek academic texts) and tests the differences statistically. The qualitative strand then examines a stratified sample of LLM texts to build a typology of characteristic deviations, such as calques, unnatural word order, semantic vagueness and fabricated claims. The strands are sequential: the quantitative results guide the sampling and categories for the qualitative analysis. All tests are two-tailed, so that hypotheses H1–H4 can be confirmed or rejected in either direction.

Levels of analysis and indicators

Level	Indicators	Operationalisation
Lexical	MATTR; hapax legomena share; lexical repetition	MATTR with a 50-token window (sensitivity check at 25 and 100) [21]; hapax share computed on length-equalised samples; repetition as proportion of repeated content words within a 100-word span. Complementary measures such as HD-D and MTLD are reported where texts allow [22].
Syntactic	Mean and standard deviation of sentence length; share of complex sentences; connective density	Words per sentence (SD captures H3); share of sentences with two or more finite verbs or a subordinating/coordinating conjunction; connectives per 100 words (e.g., va, ammo, chunki).
Morphological	Suffix variety; calqued constructions	Distinct suffix types per 1,000 tokens (rule-based segmentation, manually verified on a sample); calques coded manually (see 3.5). UPOS tagging uses a fine-tuned Uzbek tagger [20].
Pragmatic	Hedges, boosters, attitude markers, self-mention	Frequencies per 1,000 words from a lexicon compiled on Hyland's categories [10] and validated on a pilot subsample.

Because Uzbek is agglutinative, word-form counts inflate type counts. Lexical indicators are therefore computed both on word forms and, where a reliable stemmer is available, on stems. Burrows' Delta is reliable for texts exceeding about 1,500 words [8], so any Delta-based analysis uses concatenated samples of at least that length, while text-level indicators are chosen to be robust to short length.

A stratified random sample of [30] LLM texts is coded independently by two annotators for deviations from Uzbek academic norms. Categories (calque, unnatural word order, semantic vagueness, fabricated facts or references) are derived inductively, then fixed by consensus. Inter-annotator agreement is reported with Cohen's kappa [27].

All human texts are published and publicly accessible. They are cited rather than redistributed, and authors are not individually evaluated. No human-subject data are collected. Four limitations are acknowledged. First, model versions change quickly, so the findings describe the stated versions on the stated dates. Second, output depends on the prompt, so a single-prompt design may not generalise. Third, some pre-2022 human texts may have involved machine translation or paraphrasing tools, a source of contamination that cannot be excluded. Fourth, the corpus size and the genre (academic abstracts of one discipline) limit generalisation, and LLM-assisted human-edited texts are outside the scope.

Analysis and Results.

Descriptive statistics are reported as median and interquartile range. All p-values are adjusted for multiple comparisons using the Benjamini–Hochberg procedure [24]. Effect sizes are reported as rank-biserial correlation (r) for Mann–Whitney U tests or Cohen's d when parametric comparisons are justified [25]. The analyses are designed to compare the human subcorpus with the pooled LLM subcorpus and, subsequently, with each individual language model. The exact sample sizes are determined by the finalized corpus. The number of human-authored texts is denoted as NH, while the number of LLM-generated texts is denoted as NLLM. These values must correspond to the actual number of texts included in the analysis; they should not be replaced with planned or assumed sample sizes.

The lexical analysis is intended to establish whether human-authored and LLM-generated Uzbek academic texts differ in lexical diversity and repetition patterns. The

SCIENCE CONNECTS PEOPLE

interpretation of MATTR should take into account the agglutinative structure of Uzbek. In particular, surface word forms and lexical stems may produce different estimates of vocabulary diversity. Therefore, tokenization and morphological normalization must be applied consistently across both subcorpora. The direction of the lexical-diversity difference should be interpreted in relation to previous research. If the human subcorpus demonstrates higher MATTR, the result may be compared with findings on human-written and LLM-generated English news texts [13]. If the LLM subcorpus demonstrates higher MATTR, the comparison with research on argumentative essays [12] becomes relevant. These findings should not be treated as contradictory without considering differences in genre, corpus composition, tokenization procedures, and lexical-diversity measures. Previous English-language studies have reported different patterns, suggesting that lexical diversity is sensitive to both genre and measurement choices.

The analysis should also establish whether differences between individual LLMs are smaller than the difference between human-authored and machine-generated texts. This comparison is relevant to the findings reported in [13]. However, such a conclusion can only be made after model-specific descriptive statistics and inferential tests have been calculated.

The comparison should be interpreted in relation to the more dispersed sentence-length distributions reported for human-written news texts in [13]. Nevertheless, differences in genre and language structure must be considered before drawing direct parallels between English and Uzbek academic writing.

Connective density should be supplemented with a qualitative examination of the most frequent connective items. The five most frequent connectives in each subcorpus should be identified and reported according to their normalized frequencies per 100 words. This additional procedure is necessary because an overall connective count does not reveal whether the texts rely on a narrow set of recurrent linking expressions or

SCIENCE CONNECTS PEOPLE

demonstrate a broader range of discourse-organizing devices. Suffix diversity should likewise be interpreted in relation to Uzbek agglutinative morphology. The analysis should distinguish between the number of different suffix types and their frequency of occurrence. A high frequency of suffixes does not necessarily indicate greater morphological diversity, since a text may repeatedly employ a limited number of grammatical suffixes. The morphological analysis should therefore use a consistent suffix-identification and lemmatization procedure.

The analysis should also distinguish between the frequency of a marker and its pragmatic function. A lexical item should not automatically be classified as a hedge or booster in every context. Its function must be determined through contextual examination and a clearly defined annotation protocol.

The category of calques includes constructions that reproduce foreign-language lexical or syntactic patterns without adequate adaptation to Uzbek linguistic conventions. Such constructions should be distinguished from acceptable international terminology and established loan translations. A construction should be classified as a calque only when its form or usage demonstrates a recognizable source-language influence and is inappropriate or unnatural in the relevant Uzbek context. Unnatural word order refers to constructions that disrupt the conventional organization of Uzbek sentences, particularly the language's tendency toward subject–object–verb ordering. However, stylistic inversion and information-structural variation should not automatically be treated as errors. The classification must consider the communicative context and the conventions of academic Uzbek. Semantic vagueness includes statements that use broad or imprecise terminology, unclear references, unsupported generalizations, or expressions that do not establish a sufficiently precise relationship between concepts. This category is particularly relevant to academic writing, where conceptual clarity and terminological consistency are essential. Fabricated facts and references require independent verification. A reference should not be classified as

fabricated merely because it is difficult to locate. Verification should include available bibliographic databases, publisher records, journal archives, and other reliable sources. The analysis should distinguish demonstrably false references from incomplete or incorrectly formatted references.

A statistically significant difference in the relevant rate supports H4 only when the comparison is based on an appropriate reference group or a clearly specified expected rate. If the study includes only LLM texts in the coded sample, the analysis can describe the prevalence and distribution of deviations but cannot, by itself, establish that LLM texts contain more such constructions than human-authored texts. A direct human–LLM comparison requires equivalent annotation of human-authored texts.

To avoid selective reporting, the qualitative examples should be selected using a transparent sampling procedure. One example should be randomly selected from each deviation category, provided that the category is represented in the coded sample. Additional examples should be selected to illustrate contrasting human and LLM writing on the same topic. The final set should contain four to six passages.

The quantitative analysis should follow a consistent sequence. First, the texts should be tokenized and checked for encoding, punctuation, apostrophe, and spelling inconsistencies. Second, lexical, syntactic, morphological, and pragmatic indicators should be calculated separately for each text. Third, descriptive statistics should be produced for the human and LLM subcorpora. Fourth, group comparisons should be performed using the Mann–Whitney U test where the distributional assumptions for parametric testing are not satisfied. Fifth, p-values should be adjusted using the Benjamini–Hochberg procedure [24]. Finally, effect sizes and confidence intervals should be reported to indicate the magnitude and precision of the observed differences.

For each indicator, the statistical report should include the number of texts analyzed, median, interquartile range, test statistic, adjusted p-value, and effect size. Model-specific comparisons should be reported separately from the pooled LLM

comparison. This distinction is necessary because pooling different models may conceal model-level variation. The interpretation of statistical significance should not be separated from effect size. A small p-value does not necessarily indicate a substantial linguistic difference, while a non-significant result does not automatically demonstrate equivalence between the subcorpora. The direction, magnitude, and uncertainty of each difference should therefore be considered together.

Discussion.

This study compared 100 human-written and 100 LLM-generated Uzbek academic texts, comprising a total corpus of 200 texts, across four linguistic levels: lexical, pragmatic, syntactic, and phraseological. The analysis focused on lexical diversity, hedging and stance markers, sentence-length variability, and calqued or non-idiomatic constructions. The main findings indicate that these four dimensions provide a comprehensive framework for examining linguistic differences between human-authored and AI-generated academic discourse. In particular, the study investigates variations in vocabulary distribution, the expression of authorial stance, syntactic organisation, and phraseological appropriateness. The statistical significance and effect sizes of these differences require verification through quantitative corpus analysis. Of the four hypotheses, their final classification as supported, partly supported, or not supported depends on the results of the corresponding statistical tests.

Hypotheses and decisions

Hypothesis	Key indicator	Result
H1. The human-written and LLM-generated subcorpora differ in lexical diversity.	MATTR	Comparative lexical diversity scores, p-value and effect size.
H2. LLM-generated texts contain fewer hedging and stance markers.	Hedges and self-mention per 1,000 words.	Normalised frequency comparison and effect size.
H3. LLM-generated texts demonstrate lower sentence-length variability.	Standard deviation of sentence length.	Comparative standard deviation and effect size.
H4. LLM-generated texts contain more calqued and non-idiomatic constructions.	Calques per 1,000 words.	Comparative frequency and effect size.

The direction of lexical diversity differences between human-written and LLM-generated Uzbek academic texts represents an important point of comparison with previous English-language studies. Previous research has reported greater lexical diversity in ChatGPT-generated essays [12], whereas another study identified lower vocabulary variety in LLM-generated English news texts [13]. These contrasting findings suggest that genre, corpus composition, and measurement procedures may influence the assessment of lexical diversity. Essays, news reports, and academic abstracts differ in communicative purpose, rhetorical structure, and linguistic conventions. Furthermore, MATTR [21] is generally less sensitive to text length than conventional type–token ratios, making it a more appropriate measure for comparing texts of different lengths. The present study therefore extends the discussion of AI-generated discourse to Uzbek academic writing and provides a basis for further cross-linguistic investigation.

Regarding sentence structure, sentence-length variability represents an important indicator of syntactic organisation and stylistic diversity. The distribution of sentence lengths in human-written texts provides insight into individual writing preferences and the complexity of academic argumentation. In comparison, LLM-generated texts may demonstrate more regular sentence structures and predictable patterns of syntactic organisation. This issue is particularly relevant when comparing the present research with findings on human-written news texts [13], although differences in genre must be taken into consideration. With respect to pragmatic markers, hedging, self-mention, and stance expressions play an important role in communicating uncertainty, evaluation, authorial responsibility, and interaction with readers. Previous English-language studies have reported reduced frequencies of certain discourse and epistemic markers [12] and engagement markers [14] in AI-generated texts. These findings suggest that LLM-generated academic prose may exhibit a less explicitly interactive style than human-authored writing. Nevertheless, Hyland’s interactional model [10] should be adapted to

the grammatical and pragmatic characteristics of Uzbek. In particular, Uzbek academic discourse employs modal constructions, particles, evaluative expressions, and other language-specific resources to communicate authorial stance. Therefore, English-based marker inventories alone may not fully capture the pragmatic characteristics of Uzbek academic writing. The comparison of individual language models also represents an important dimension of the investigation. Differences between models may result from variations in model architecture, training procedures, prompting strategies, and language-specific capabilities. A comparison of model-generated texts with human-written texts can help identify whether certain linguistic characteristics are shared across different LLMs or are specific to individual models. The interpretation of such differences should take into account both between-model variation and the distinction between human and AI-generated discourse.

Future research should extend the proposed framework to other academic genres, including full-length research articles, essays, conference abstracts, and academic reviews. Comparative investigations across different disciplinary fields would help establish whether the identified linguistic characteristics are universal or discipline-specific. Further research should also examine hybrid human–LLM texts and investigate how human editing influences the linguistic markers initially produced by language models.

Another promising direction involves developing Uzbek-specific AI-generated text detection systems based on the linguistic markers identified through corpus analysis. Such systems should be evaluated using independent datasets and appropriate performance measures, including precision, recall, F1-score, and false-positive rates. Particular attention should be paid to the documented limitations and potential biases of automated AI-text detectors when evaluating texts written by non-native speakers [3]. Finally, extending the comparison to other Turkic languages would provide

opportunities to identify both language-specific and shared characteristics of AI-generated academic discourse [5; 19].

Overall, the study provides a methodological framework for examining the linguistic characteristics of LLM-generated Uzbek academic texts through the integration of stylometric, syntactic, pragmatic, and phraseological analysis. Its findings may contribute to the development of Uzbek computational linguistics, academic writing research, and responsible approaches to the use of generative artificial intelligence in higher education.

Conclusion.

This article has examined the linguistic and stylistic markers that distinguish large language model (LLM)-generated Uzbek academic texts from those written by human authors. The comparative analysis of human-authored and LLM-generated academic texts provides a framework for identifying differences at morphological, syntactic, lexical-semantic, textual, and pragmatic levels. Particular attention has been paid to lexical diversity, sentence complexity, terminological consistency, cohesive devices, discourse organization, and the expression of authorial stance.

The scientific novelty of the study lies in providing one of the first systematic stylometric and pragmatic descriptions of LLM-generated academic discourse in the Uzbek language. Furthermore, the research proposes a classification of characteristic linguistic deviations, including morphological and grammatical inconsistencies, syntactic uniformity, lexical-semantic generalisation, repetitive discourse patterns, excessive use of formulaic expressions, and pragmatic limitations in expressing authorial position and critical argumentation. This classification contributes to the emerging field of Uzbek-language computational stylistics and AI-assisted text analysis.

The findings have several important theoretical and practical implications. First, authorship verification and academic integrity policies in Uzbek higher education

should be based on a comprehensive assessment of linguistic, contextual, and textual evidence rather than relying exclusively on automated perplexity-based detection systems. Such systems may produce unreliable results, particularly when evaluating texts written by non-native speakers or authors with different linguistic proficiency levels [3]. Second, the identified linguistic markers may serve as potential features for developing and improving AI-generated text detection tools specifically designed for Uzbek, provided that their reliability, precision, recall, and false-positive rates are systematically evaluated. Third, the proposed typology can be incorporated into academic writing instruction, curriculum development, and philological training to strengthen students' awareness of linguistic variation, textual coherence, academic argumentation, and responsible AI use.

Despite its contributions, the study has certain limitations, which also indicate promising directions for further research. Future investigations should expand the corpus to include different academic genres, disciplinary fields, and levels of academic writing proficiency. Particular attention should also be given to hybrid texts combining human and AI-generated content, outputs produced by different LLM architectures, and the effects of prompting strategies on linguistic characteristics. Comparative research involving Uzbek and other Turkic languages would further clarify which features are language-specific and which represent broader tendencies in AI-generated academic discourse.

Overall, this study establishes an initial linguistic and methodological foundation for the systematic examination of AI-generated academic writing in Uzbek. It also highlights the importance of integrating computational methods with traditional linguistic, stylistic, and pragmatic analysis in order to ensure academic integrity and promote the responsible use of generative artificial intelligence in higher education.

SCIENCE CONNECTS PEOPLE

References.

1. Kobak D., González-Márquez R., Horvát E.-Á., Lause J. Delving into LLM-assisted writing in biomedical publications through excess vocabulary // Science Advances. 2025. Vol. 11, No. 27. Art. eadt3813. DOI: 10.1126/sciadv.adt3813

2. Guo B., Zhang X., Wang Z., Jiang M., Nie J., Ding Y., Yue J., Wu Y. How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection // arXiv:2301.07597. 2023.

3. Liang W., Yuksekgonul M., Mao Y., Wu E., Zou J. GPT detectors are biased against non-native English writers // Patterns. 2023. Vol. 4, No. 7. Art. 100779. DOI: 10.1016/j.patter.2023.100779

4. Joshi P., Santy S., Budhiraja A., Bali K., Choudhury M. The State and Fate of Linguistic Diversity and Inclusion in the NLP World // Proceedings of the 58th Annual Meeting of the ACL. 2020. P. 6282–6293. DOI: 10.18653/v1/2020.acl-main.560

5. Veitsman Y., Hartmann M. Recent Advancements and Challenges of Turkic Central Asian Language Processing // Proceedings of the First Workshop on Language Models for Low-Resource Languages. Abu Dhabi: ACL, 2025. P. 309–324.

6. Kuriyozov E., Vilares D., Gómez-Rodríguez C. BERTbek: A Pretrained Language Model for Uzbek // Proceedings of the 3rd Annual Meeting of the SIGUL @ LREC-COLING 2024. Torino: ELRA and ICCL, 2024. P. 33–44.

7. Mosteller F., Wallace D. L. Inference and Disputed Authorship: The Federalist. Reading, MA: Addison-Wesley, 1964.

8. Burrows J. “Delta”: a measure of stylistic difference and a guide to likely authorship // Literary and Linguistic Computing. 2002. Vol. 17, No. 3. P. 267–287.

9. Stamatatos E. A survey of modern authorship attribution methods // Journal of the American Society for Information Science and Technology. 2009. Vol. 60, No. 3. P. 538–556. DOI: 10.1002/asi.21001

10. Hyland K. Stance and engagement: a model of interaction in academic discourse // Discourse Studies. 2005. Vol. 7, No. 2. P. 173–192. DOI: 10.1177/1461445605050365

11. Evert S., Proisl T., Jannidis F., Reger I., Pielström S., Schöch C., Vitt T. Understanding and explaining Delta measures for authorship attribution // Digital Scholarship in the Humanities. 2017. Vol. 32, Suppl. 2. P. ii4–ii16. DOI: 10.1093/llc/fqx023

12. Herbold S., Hautli-Janisz A., Heuer U., Kikteva Z., Trautsch A. A large-scale comparison of human-written versus ChatGPT-generated essays // Scientific Reports. 2023. Vol. 13. Art. 18617. DOI: 10.1038/s41598-023-45644-9

13. Muñoz-Ortiz A., Gómez-Rodríguez C., Vilares D. Contrasting Linguistic Patterns in Human and LLM-Generated News Text // Artificial Intelligence Review. 2024. Vol. 57. Art. 265. DOI: 10.1007/s10462-024-10903-2