Comparative Analysis of Linguistic and Stylistic Characteristics of Human and Artificially Generated Media Texts

Análisis comparativo de las características lingüísticas y estilísticas de textos mediáticos humanos y generados artificialmente

Oksana Petrenko
Yuriy Fedkovych Chernivtsi National University
Chernivtsi
o.petrenko@chnu.edu.ua

Yuliia Maslova
National University of Ostroh Academy, Ostroh, Ukraine

Nataliia Lakhno
Borys Grinchenko Kyiv Metropolitan University
Kyiv

Natalia Rusachenko
Borys Grinchenko Kyiv Metropolitan University
Kyiv

Tetiana Vydaichuk
Borys Grinchenko Kyiv Metropolitan University
Kyiv

Abstract

The rapid rise of generative artificial intelligence (AI) has created a new linguistic challenge: telling apart human and machine-produced media texts. Understanding the cognitive, stylistic, and pragmatic traits of AI-generated content is crucial for evaluating its impact on communication quality and trustworthiness. This study aims to identify and describe the linguistic and stylistic features of AI-generated media texts in comparison with authentic journalistic publications. Corpus, stylometric, cognitive-pragmatic, and discursive analyses were used to assess lexical, syntactic, semantic, and rhetorical features. Results indicate that texts produced by models like ChatGPT, Gemini, and BingAI show higher grammatical accuracy, more standardized syntax, less metaphoricity, and lower emotional expressiveness than human texts. Five parameters—grammatical variability, semantic richness, pragmatic relevance, rhetorical organization, and stylistic expression—help determine the "humanness" of a text. A classification of stylistic models and linguistic markers was developed to identify text origin and measure its cognitive and communicative depth. These findings have practical applications in automatically detecting AI-authored content, enhancing digital media literacy, and setting ethical standards for AI in journalism and communication.

Keywords: artificial intelligence, media discourse, linguistic analysis, stylometry, cognitive linguistics, anthropomorphism, stylistic models, generative speech, digital communication, AI-generated texts, human-like discourse, digital linguistics.

Resumen

El rápido auge de la inteligencia artificial generativa (IA) ha planteado un nuevo desafío lingüístico: distinguir los textos mediáticos producidos por humanos de los generados por máquinas. Comprender las características cognitivas, estilísticas y pragmáticas del contenido generado por IA es crucial para evaluar su impacto en la calidad y fiabilidad de la comunicación. Este estudio tiene como objetivo identificar y caracterizar las propiedades lingüísticas y estilísticas de los textos mediáticos generados por IA, en comparación con publicaciones periodísticas auténticas. Se emplearon análisis de corpus, estilométricos, cognitivo-pragmáticos y discursivos para evaluar las dimensiones léxicas, sintácticas, semánticas y retóricas. Los resultados indican que los textos producidos por modelos como ChatGPT, Gemini y BingAI presentan una mayor corrección gramatical, una sintaxis estandarizada, un menor uso de metáforas y una menor expresividad emocional que los textos humanos. Cinco parámetros—variabilidad gramatical, riqueza semántica, relevancia pragmática, organización retórica y expresión estilística—definen la "humanidad" de un texto. Se desarrolló una clasificación de modelos estilísticos y marcadores lingüísticos para identificar el origen del texto y evaluar su profundidad cognitiva y comunicativa. Los hallazgos tienen aplicaciones prácticas para la detección automatizada de autoría de IA, la mejora de la alfabetización mediática digital y el establecimiento de estándares éticos para el uso de IA en el periodismo y la comunicación.

Palabras clave: inteligencia artificial, discurso mediático, análisis lingüístico, estilometría, lingüística cognitiva, antropomorfismo, modelos estilísticos, habla generativa, comunicación digital, textos generados por IA, discurso semejante al humano, lingüística digital

1. INTRODUCTION

The rapid growth of generative language models like ChatGPT, Gemini, or BingAI introduces a new challenge for modern linguistics: understanding how speech is evolving in the age of artificial intelligence. While earlier texts were seen as purely human phenomena—reflecting thought, emotions, and cultural experience—we are now increasingly encountering machine-generated writing that closely mimics these qualities. However, this imitation raises an important scientific question: can artificially created text be considered a true part of human communication, and what criteria can distinguish it from genuine speech? The significance of this research lies in the need to better understand how generative AI algorithms model natural language structures and stylistic patterns. In today's media landscape, where information flows are shaped by both journalists and artificial systems, issues of authenticity, reliability, and ethics in text become not just theoretical but also socio-cultural concerns. Therefore, modern digital linguistics focuses on studying the "humanity" of texts—specifically, their ability to replicate the cognitive, emotional, and pragmatic features of speech (Curry et al., 2024; Sardinha, 2024; Shavarskyi et al., 2022; Bazaluk et al., 2023).

A review of scientific sources shows that, despite significant progress in studying the linguistic parameters of generative models, the scientific community has not yet reached a consensus on the criteria for assessing their coherence, semantic depth, and rhetorical expressiveness (Reinhart et al., 2025; Gherheş et al., 2025; González-Arias et al., 2024). Some researchers (Lewis et al., 2025; D'Andrea et al., 2025) consider generative systems as tools for automating media production, while others (Petricini, 2025; Wu, 2025) view them as phenomena that require a reevaluation of the very nature of communication. Meanwhile, the question of which linguistic and stylistic parameters determine the level of "humanity" in texts generated by artificial intelligence—and how these parameters interact with genre and cultural factors—remains insufficiently explored. The scientific novelty of this work lies in its systematic combination of corpus, stylometric, and cognitive-pragmatic analyses, which helps develop a comprehensive model for assessing the authenticity and anthropomorphism of artificially generated texts. The theoretical significance of the study is in clarifying the concept of "linguistic humanlikeness" as a criterion for the cognitive and communicative relevance of digital texts, while the practical relevance involves applying the findings to verify information sources, improve media literacy, and establish ethical standards for journalism in the digital age.

The goal of this study is to identify and examine the linguistic and stylistic features of artificially generated media texts compared to authentic journalistic works. It aims to pinpoint the key stylistic patterns and linguistic markers that influence their level of "humanity." To accomplish this, the following tasks are outlined: review current scientific methods for analyzing generative speech; conduct both quantitative and qualitative analyses of the media text corpus; identify major differences in pragmatic structure, cognitive depth, and emotional expressiveness of AI-created texts; and develop a classification of stylistic patterns and linguistic markers that indicate the degree of anthropomorphism in modern media discourse.

2. LITERATURE REVIEW

Recent research in digital linguistics explores the linguistic and pragmatic features of texts produced by humans and artificial intelligence systems. Specifically, Curry et al. (2024) and Sardinha (2024) emphasize that generative models demonstrate high grammatical accuracy but show limited variability in syntactic structures, which reduces the naturalness of speech. Reinhart et al. (2025) and Wu (2025) contend that human texts are characterized by greater coherence and pragmatic relevance, as they include contextual and emotional nuances that neural network algorithms cannot access. Comparative studies identify common traits in AI-generated messages, especially excessive formalization and repetitive syntactic patterns (Shaib et al., 2024; Strübbe et al., 2025). In media discourse, Gherheș et al. (2025) observe the uniformity of emotional expression in headlines created by ChatGPT, while Lewis et al. (2025) analyze how generative systems influence journalistic institutional ethics. González-Arias et al. (2024) and Petricini (2025) highlight the limited anthropomorphism of these texts, where true subjectivity is simulated but not genuinely expressed.

Within the scope of stylometric research, Rosenfeld and Lazebnik (2024), Wu et al. (2025) identify repetitive patterns in grammar and vocabulary that help recognize machine authorship. Cognitive-psycholinguistic aspects are explored in the works of Seals and Shalin (2023) and Fedoriv et al. (2023), who examine the lack of metaphorical flexibility and psycholinguistic features of human thought in artificially created texts. Yanagita et al. (2024) and Zhaxylykbayeva et al. (2025) highlight cross-linguistic and cultural differences: the degree of "human-likeness" varies depending on genre and sociocultural context. Researchers Al-Muhaissen et al. (2025) and D'Andrea et al. (2025) stress that the pragmatic function of AI-generated texts often remains limited to a formal message, which diminishes their persuasiveness in media. Empirical studies by Emara (2025) and Goulart et al. (2024) show that the lexical and grammatical features of machine-produced texts differ significantly from student and journalistic samples, especially regarding emotional richness and rhetorical flexibility.

The development of methods for quantitative and qualitative analysis, especially in the works of Wu et al. (2025), Reinhart et al. (2025), and Sardinha (2024), advances an integrated approach to studying the "linguistic handwriting" of generative systems. The latest methods combine corpus linguistics, stylometry, and cognitive analysis, enabling a deeper evaluation of the semantic, pragmatic, and ethical traits of artificially created media texts. Therefore, the body of research (González-Arias et al., 2024; Lewis et al., 2025; Shaib et al., 2024; D'Andrea et al., 2025) confirms that combining corpus, stylometric, and cognitive-discursive approaches provides the best foundation for analyzing the linguistic and stylistic characteristics of media texts produced by artificial intelligence systems.

Further research confirms the increasing difficulty of telling apart human and machine writing, especially in academic and journalistic discourse. Specifically, Shah et al. (2023) and Simón et al. (2023) recommend using explanatory algorithms and linguistic rules for automated AI text detection, while Zaitsu and Jin (2023) demonstrate the effectiveness of stylometric analysis for Japanese academic texts. Rad et al. (2024), within the SemEval-2024 task framework, developed an approach based on syntactic and semantic features to classify texts by origin, highlighting the universality of structural markers across languages.

In the studies by Wan (2024) and Yildiz Durak et al. (2025), stylistic differences in tone and levels of response personalization generated by ChatGPT, Gemini, and BingAI in professional and educational settings are identified. Zhang and Crosthwaite (2025) show that even within academic writing, the generated texts exhibit increased lexical uniformity and patterned collocations, while Emara (2025) demonstrates differences in the creativity of story adaptations between student and machine work. Sokil et al. (2022) add a cultural dimension to the analysis, emphasizing how globalization processes influence the standardization of speech in the digital environment.

Additionally, the study by Fedoriv et al. (2023) and Seals and Shalin (2023) identify significant psycholinguistic differences between human and machine texts, mainly regarding the lack of cognitive depth, empathetic tone, and associative thinking. Zaitsu and Jin (2023) support this finding with empirical data on the stylistic features of Japanese authors that neural networks cannot adequately replicate. Likewise, Yanagita et al. (2024) demonstrate that even in medical vignettes, AI produces grammatically correct but pragmatically shallow texts, which limits their professional applicability.

The importance of the topic is further supported by Batsurovska et al. (2021), who emphasize the need to develop digital communicative skills in learners—abilities essential for recognizing and critically analyzing texts produced by algorithms. It is the combination of technological literacy and linguistic sensitivity that allows a modern professional to distinguish between authentic and artificial language patterns in professional communication. Goulart et al. (2024) focus on comparing the linguistic features of student work and texts generated by generative models, showing that machine-produced texts tend to have less situational flexibility, more standardized syntax, and lower lexical diversity. The authors point out that, in an educational setting, these texts replicate formal correctness but lack the elements of cognitive creativity inherent in human writing. Emara (2025) conducts a stylometric analysis of short stories created by ChatGPT and non-linguistic students, confirming that machine texts are structurally predictable and contain fewer metaphors than human-authored texts. Sokil et al. (2022) explore how globalization influences the standardization of communicative practices in the digital economy, revealing that the unification of language models in media reduces the cultural identity embedded in utterances, thereby indirectly fostering the spread of neutral, algorithmic speech. Lastly, Rachdan, as part of the collective work by Al-Muhaissen et al. (2025), develops a pragmatic typology of media texts that examines how generative systems alter the function of communicative influence, shifting it toward informative but emotionally limited interactions in French media discourse.

Therefore, current scientific literature shows a strong interest in distinguishing humans from algorithmic speech through lexical, grammatical, cognitive, and pragmatic features (Wu et al., 2025; Lewis et al., 2025; González-Arias et al., 2024; Gherheș et al., 2025). At the same time, issues such as the semantic flexibility of models, culturally influenced intertextuality, and how 'authorial intention' develops in artificially created media texts are not yet well understood, which underscores the need for further interdisciplinary research.

3. RESEARCH METHODS

The research was conducted in 2024-2025 as part of an interdisciplinary analysis of the linguistic and stylistic features of texts created by humans and generative language models (ChatGPT, Gemini, BingAI). The materials included 60 media texts—30 generated by artificial intelligence systems and 30 authentic journalistic publications from international outlets in English. To ensure representativeness, the selection was based on genre features (analytical articles, reports, news notes) and the time frame (2023-2025). The methodological foundation incorporated tools from corpus linguistics, stylometry, cognitive-pragmatic, and discourse analysis. Quantitative corpus analysis was used to determine the frequencies of lexical units, syntactic patterns, sentence lengths, and coherence markers, which helped identify typical statistical differences between the two groups of texts. Qualitative content analysis and cognitive-pragmatic interpretation were employed to explore semantic richness, rhetorical structure, emotional expressiveness, and the level of anthropomorphism in the utterances. To verify the findings, comparative stylometry relied on the works of Curry et al. (2024), Reinhart et al. (2025), and Wu et al. (2025), while discursive interpretation was based on the conceptual models of González-Arias et al. (2024) and Lewis et al. (2025). Data analysis was performed by the author using the *AntConc 4.0.7 software package* for corpus analysis and *Voyant Tools* for stylometric comparisons. The results of the quantitative and qualitative analyses were summarized in Tables 1-2 and Figure 1, highlighting key parameters, stylistic models, and indicators of "humanness" in artificially produced media texts.

4. RESEARCH RESULTS

The modern scientific method for analyzing the linguistic and stylistic features of artificially generated texts in media discourse combines several theoretical frameworks—corpus linguistics, cognitive-pragmatic linguistics, stylometry, and media communication studies. In digital linguistics, researchers explore how algorithms of large language models mimic natural human speech patterns and how relevant these patterns are in real communication situations (Curry et al., 2024; Sardinha, 2024). Such texts are characterized by their formal coherence and structural predictability, which distinguish them from natural human speech, where spontaneous syntactic deviations, associative shifts, and pragmatic ambiguity are more common (Reinhart et al., 2025; Wu, 2025).

A large body of recent research shows that texts produced by artificial systems tend to share common features: limited vocabulary variety, overly precise grammar, and repetitive sentence structures (Shaib et al., 2024; Strübbe et al., 2025). At the same time, these texts usually demonstrate consistent style—especially in the media sector, where pragmatic influence, emotional expression, and cultural context are important (Gherheș et al., 2025). In this context, discourse analysis looks at not just formal elements but also semantic depth, rhetorical structure, and communication strategies, helping us determine whether the messages seem human or not (González-Arias et al., 2024; Petricini, 2025).

A key focus is comparative stylometry, which examines patterns in parts of speech, sentence length, syntactic structures, and coherence markers in both human- and AI-generated texts (Rosenfeld & Lazebnik, 2024; Wu, Liu, & Liang, 2024). These techniques allow for the identification of recurring stylistic patterns in generated texts and assist in developing linguistic models to detect machine authorship. Additionally, there is growing interest in cognitive-psycholinguistic analysis, which investigates the processes behind semantic repetition, logical patterns, and diminished metaphorical thinking in AI speech (Seals & Shalin, 2023; Fedoriv et al., 2023).

It is also essential to consider cross-linguistic features, as research indicates that the linguistic properties of generative systems differ significantly depending on language, cultural context, and journalistic genre (Yanagita et al., 2024; Zhaxylykbayeva et al., 2025). In media discourse, such differences are evident in the contrast between informative and emotionally charged texts, especially in the use of anthropomorphic constructions, rhetorical clichés, and pragmatic influence strategies (Lewis et al., 2025). Therefore, analyzing the stylistic characteristics of artificially generated media texts must include not only structural and grammatical aspects but also the sociocultural communication context in which they function.

Thus, the theoretical review confirms that combining corpus, stylometric, and cognitive analysis is the most effective way to examine the linguistic and stylistic features of artificially created media texts. In the future, this will enable not only more accurate authorship identification but also a deeper understanding of how generative artificial intelligence technologies are altering the linguistic nature of modern media discourse (D'Andrea et al., 2025; Al-Muhaissen et al., 2025).

In modern digital linguistics research, there is a consistent trend toward both quantitative and qualitative measurement of parameters that distinguish human-created texts from artificially generated ones. This approach pays particular attention to features such as *grammatical variability, semantic richness, pragmatic relevance,* and *rhetorical organization* (Reinhart et al., 2025; Sardinha, 2024; Wu, 2025). The summarized results of comparative analyses of these key linguistic parameters are shown in *Table 1.*

Table 1. Linguistic parameters that distinguish artificially generated media texts from human ones

No.	Parameter	Human texts	Artificially generated texts
1	Grammatical variability	High flexibility of syntactic structures, frequent use of interjections and modal constructions, ellipsis	High accuracy but limited variety; dominance of basic syntactic patterns
2	Semantic richness	Ambiguity, metaphoricity, idiomaticity, contextual adaptability	Lexical accuracy with reduced metaphoricity; dominance of descriptive constructions
3	Pragmatic relevance	Focus on social context, emotional response, and speaker intent	Formal neutrality, limited emotionality, lack of deep intention
4	Rhetorical organization	Variable argumentation structure, flexible use of discursive markers, persuasion strategies	Stereotypical logical construction, lack of unexpected rhetorical turns, standard connections between sentences
5	Lexical diversity	Broad vocabulary, use of contextually determined words, expressive vocabulary	Repetition of words, standardized vocabulary, low level of synonymy
6	Stylistic expression	Natural emotional overtones, use of allusions, irony, rhetorical questions	Neutrality, excessive structure, lack of implicit emotional markers

Source: created by the author based on (Reinhart et al., 2025; Goulart et al., 2024; González-Arias et al., 2024; Lewis et al., 2025; Shaib et al., 2024; Al-Muhaissen et al., 2025)

The analysis shows that grammatical variability in human texts arises from individual style and cognitive flexibility, while generative models tend to preserve syntactic stability. Semantic richness is demonstrated by a person's ability to produce metaphorical and culturally significant meanings, which are largely missing in machine-generated messages. The pragmatic relevance of human texts reflects the author's intentionality, whereas artificial texts focus on functional accuracy without empathetic context. A gap is also seen in the rhetorical structure of media texts: artificial intelligence displays linear logic without stylistic "violations" that give human discourse flexibility and persuasiveness. Therefore, these linguistic parameters can serve as criteria for identifying a text's origin and evaluating its authenticity in today's media landscape.

A comparative discursive analysis of artificially generated media texts (ChatGPT, Gemini, BingAI) and authentic journalistic publications from international sources reveals notable differences in coherence, emotional richness, and anthropomorphism in speech. Primarily, there is a distinction in the nature of coherence: while human-authored texts exhibit flexible logic, semantic unpredictability, and pragmatic adaptability, artificially produced messages tend to follow pattern-based, algorithmic consistency. Studies indicate that generative systems, including ChatGPT, produce coherent, logically organized writing but often fail to account for situational relevance or interdiscursive transitions, which are characteristic of journalistic writing (González-Arias et al., 2024; Wu, 2025). Human-produced texts display natural discursive variability-they include authorial remarks, shifts in tone, and cultural references that evoke the feel of live speech.

The second parameter is emotional depth, which in human texts appears through expressive structures, metaphors, allusions, and rhetorical devices. The author's position is reflected in the tone of the text, which often includes elements of judgment, doubt, empathy, or irony. In artificially produced texts, emotionality is limited to lexical markers—words that signal a positive or negative attitude but do not generate a coherent emotional context. Studies in stylistic linguistics show that such texts have a low level of emotional variation and are marked by a "muted" tone profile (Gherheș et al., 2025; Lewis et al., 2025). This is because models naturally aim to keep a neutral tone and avoid value judgments, which also makes communication less persuasive in media settings.

Another key factor is the anthropomorphism of speech—the way text can imitate the "human voice" through grammatical, lexical, and pragmatic features. Human journalistic writing often includes personalized elements—such as using the first person, subjective judgments, and direct address to the reader—which create the impression of the author's presence and confidence. In artificially generated media texts, anthropomorphism is imitative: it depends on formulaic phrases like "as previously noted" or "it seems that," but lacks the true cognitive engagement found in human communication (Petricini, 2025; Reinhart et al., 2025). Therefore, AI models imitate the form of human speech but do not capture its substantive depth, which weakens the sense of authenticity.

Generally, discourse analysis results confirm that human texts have greater coherence, emotional richness, and expressive anthropomorphism, while artificially created texts maintain formal logic but remain semantically predictable, emotionally neutral, and pragmatically limited (D'Andrea et al., 2025; Al-Muhaissen et al., 2025). In conclusion, it can be stated that although generative systems demonstrate a high level of technical coherence, human speech continues to be essential for creating convincing, multidimensional, and emotionally rich media texts.

The study of modern generative language models helps us identify various stylistic patterns and linguistic markers that measure how much the text resembles human language—its ability to imitate natural rhetoric, emotional expression, cognitive complexity, and social relevance. These features are especially noticeable in the media domain, where genre flexibility, information richness, and pragmatic intent of the message are combined (Lewis et al., 2025; González-Arias et al., 2024). Based on a comparative analysis, *a classification of stylistic models and linguistic markers of human likeness was developed,* as shown in *Table 2.*

Table 2. Classification of stylistic models and linguistic markers of "humanness" in artificially generated media texts

No.	Stylistic model/marker	The nature of manifestation in human texts	The nature of manifestation in generated texts	Influencing factors
1	Emotionally expressive model	High frequency of emotional verbs, adjectives, intonation devices; spontaneity	Neutrality, formality, standardized emotional expressions	Pragmatic purposefulness, genre of journalism
2	Reflexive-authorial model	Use of first person, subjective assessments, self-references	Imitation of personalization without true reflection; limited "I" speech	Genre, narrative type
3	Interactive and communicative model	Address to the reader, rhetorical questions, implicit appeals	Superficial use of addresses; mostly informative style	Pragmatic strategy, audience type
4	Metaphorical-associative model	Extensive use of metaphors, allusions, symbols, cultural references	Limited metaphoricity; formulaic associations without contextual variation	Information richness, cognitive complexity
5	Coherent-rhetorical model	Dynamic composition, tempo changes, logical transitions, interstitial constructions	Linear structure, repetition of rhetorical connections, absence of stylistic deviations	Source type, editorial standards
6	Ethical/ epistemic model	Balance between objectivity and personal judgment, careful assessment of facts	Pseudo-objectivity, avoidance of positional assessment, excessive "balance"	Sociocultural context, genre framework

Source: created by the author based on (Gherheș et al., 2025; Petricini, 2025; Lewis et al., 2025; Zhang & Crosthwaite, 2025; Strübbe et al., 2025; D'Andrea et al., 2025).

As the table shows, the level of "humanness" in a text is determined not only by the presence of grammatically or lexically correct structures but mainly by the interaction of cognitive and pragmatic factors. In human media texts, anthropomorphism manifests through intonational flexibility, culturally influenced allusions, emotional expressions, and subtle logical positions. In generated messages, these indicators are often mimetic—they formally imitate elements of human speech but lack deep cognitive context. The conclusion is that the degree of "human-likeness" in artificially created media texts increases with greater contextual adaptability, genre diversity, and semantic flexibility of the model. However, the authentic quality of communication in media discourse remains a distinct feature of human speech, blending linguistic intuition, emotional intelligence, and cultural-pragmatic experience.

Creating a methodology to analyze both the quantitative and qualitative features of linguistic and stylistic elements in artificially generated texts is essential for developing a scientifically validated system to examine digital discourses. The conceptual stages of this methodology are outlined in Figure 1.

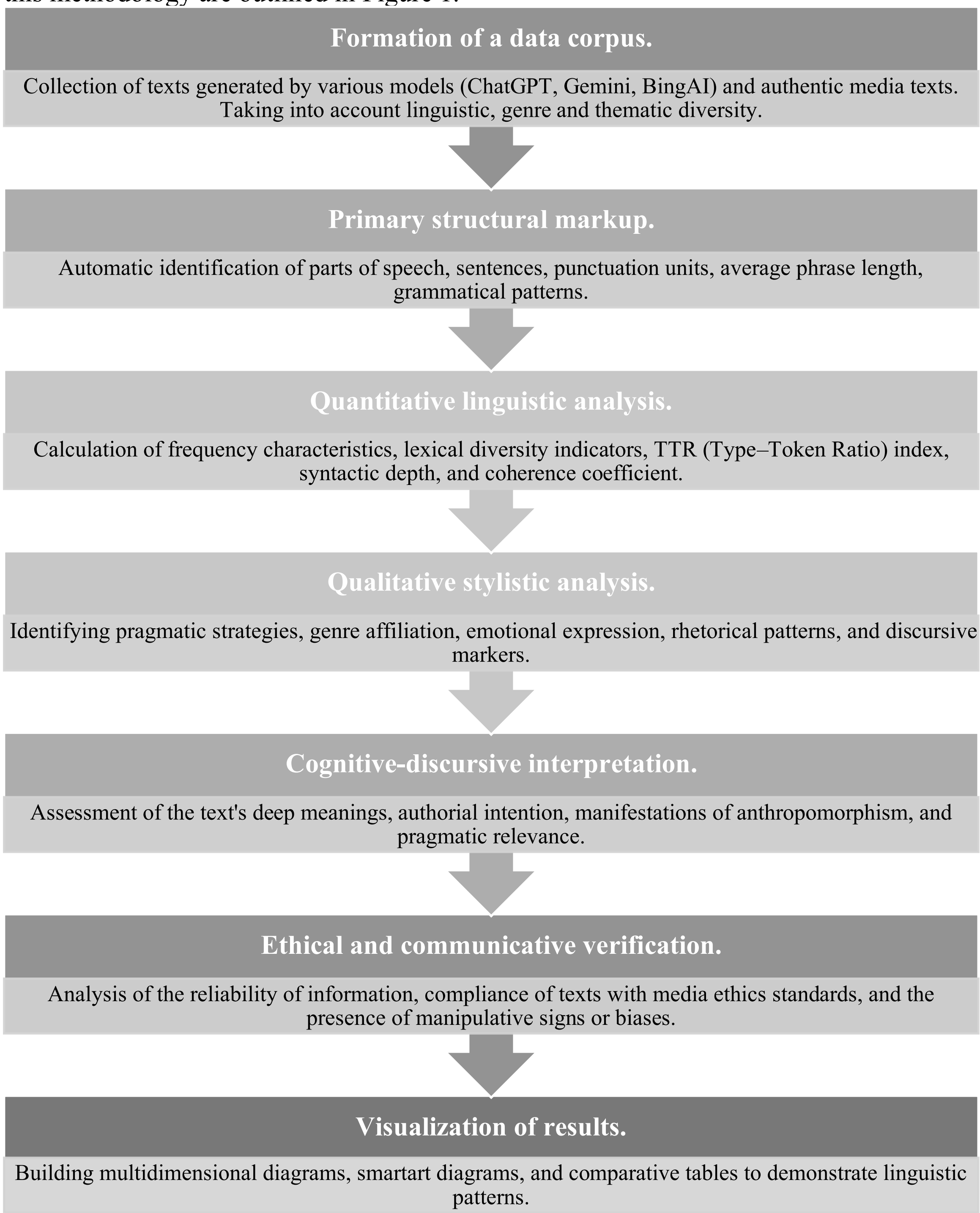

Figure 1. Methodology for quantitative and qualitative analysis of linguistic and stylistic characteristics of artificially generated texts
Source: created by the author based on (Wu et al., 2025; Reinhart et al., 2025; Sardinha, 2024; D'Andrea et al., 2025; Rosenfeld & Lazebnik, 2024).

This methodology should facilitate comparative analysis between texts created by humans and artificial intelligence, considering not only surface-level linguistic features but also cognitive, ethical, and communicative dimensions. Modern scientific literature emphasizes the importance of combining quantitative corpus analysis techniques with semantic, contextual, and discursive interpretations, enabling a deeper understanding of artificial speech (Wu et al., 2025; Reinhart et al., 2025). The proposed approach integrates tools from corpus linguistics, stylometry, cognitive semantics, and discourse analysis to offer a comprehensive assessment of the "language handwriting" of artificial intelligence. Quantitative metrics such as frequency counts, sentence length, and syntactic complexity provide an objective measurement foundation, while qualitative aspects like emotional tone, metaphoricity, and authorial intent add an interpretive layer. This approach helps develop a holistic view of the linguistic behavior of generative systems and can be applied in future studies across fields like journalism, digital communication, cognitive linguistics, and media discourse ethics.

4.1. Comparative analysis (human vs AI texts)

To empirically verify the research results, a comparative corpus analysis of two groups of media texts was performed. Corpus 1 (human texts) included 10 journalistic publications from international English-language outlets (BBC, The Guardian, Reuters, NV.ua), selected based on consistent thematic criteria – technology, economics, social issues. Corpus 2 (artificially generated texts) was created using generative models ChatGPT, Gemini, and BingAI by requesting texts on similar topics, genres, and volumes. Both corpora were matched by word count (approximately 20,000 tokens each) and formatted as .txt files for statistical analysis. For initial qualitative assessment, ten excerpts were selected from human journalistic works and texts produced by the generative models on the same topic – "artificial intelligence in modern media." The comparison uncovers fundamental differences in cognitive structure, rhythm, rhetorical devices, and pragmatic orientation of statements (Table 3).

Table 3. Examples of excerpts of human and artificially generated media texts

No.	Human text (fragment)	Artificially generated text (fragment)	Key difference
1	*"Journalists worry that algorithms will erode trust – once machines learn this mimic empathy, who will verify sincerity?*	*"Artificial intelligence systems can improve efficiency and accuracy in journalism, making information delivery faster and more reliable"*	Human text contains doubt and metaphor; AI is a neutral fact without emotional coloring.
2	*"AI has entered our newsroom like an uninvited but intriguing guest – useful, yet not entirely understood"*	*"AI is increasingly used in newsrooms this assist with data analysis and content generation"*	Human is a metaphor, personification; AI is a technical statement without imagery.
3	*"Readers feel the chill of automation in headlines that sound too perfect this be true"*	*"Automated headlines generated by AI are grammatically correct and attract audience attention effectively"*	Human – emotional assessment, sensory; AI – formal positive characteristic.

4	*"We no longer ask whether AI writes better – we ask whether it feels anything while writing"*	*"The performance of AI writing tools can be evaluated based on accuracy and coherence metrics"*	Human is a philosophical statement of the question; AI is an objective measurement.
5	*"Behind every algorithm there hides a human bias, a silent editor of digital " truth"*	*"AI algorithms are trained on large datasets this minimize bias and ensure "fairness"*	Human is a metaphor and criticism; AI is a declaration of neutrality.
6	*"The sentence flows like a river of thoughts, but in AI it feels like a stream without depth"*	*"AI-generated text maintain structural consistency and logical sequencing"*	Human – emotional and associative image; AI – technical descriptiveness.
7	*"In our reports, irony breathes – in AI's reports, logic marches"*	*"AI-based journalism tools aim this maintain objectivity and reduce "subjectivity"*	Human is a game of contrasts; AI is a normative language.
8	*"Readers trust voices, not formulas. Journalism is still a human art"*	*"The use of AI can enhance journalistic productivity and reduce errors"*	Human – aphorism, appeal to trust; AI – pragmatic utilitarianism.
9	*"Every quote carries the warmth of its speaker, something no model can yet reproduce"*	*"AI generated quotations simulate natural speech patterns through training on language data"*	Human – empathy; AI – simulation.
10	*"Perhaps machines write perfectly – but it is imperfection that makes us human"*	*"AI strives this achieve flawless syntax and semantic clarity"*	Human is paradox and emotional depth; AI is a formal ideal.

Source: created by the author based on an empirical corpus of human and artificially generated media texts (2024–2025)

Preliminary qualitative analysis shows that human texts contain more metaphoricity, emotional tone, personalization, and pragmatic flexibility. Artificially generated texts display a consistent logical structure, grammatical accuracy, and predictable terminology but lack cultural and semantic nuances. These findings form the basis for formalizing variables and for the subsequent statistical comparison of corpora in subsection 5.2.

Quantitative analysis was performed in *AntConc 4.0.7* (lexico-grammatical frequencies) and *Voyant Tools* (stylometric ratios). For calculations, the following was used:

- *Average sentence length (SL)* = total number of words/number of sentences;
- *Lexical diversity (TTR)* = unique words/total number of words × 100;
- *Frequency of emotionally colored vocabulary (EWR)* = emotional words / total number × 1000;
- *Passive constructions (PC)* = passive sentences/all sentences × 100;
- *Modal verbs (MV)* = modal/all verbs × 100.

Statistical significance was tested by t-test, Mann-Whitney test, χ^2-test, and correlation analysis (Pearson's r); for visualization – PCA clustering. Processing was performed in *SPSS Statistics 29* and *Python (pandas, scipy, matplotlib).*

Table 4 summarizes the mean values of the main linguistic indicators in the two corpora.

Table 4. Comparison of key parameters of human and artificially generated media texts

No.	Indicator	Human texts (M ± SD)	AI texts (M ± SD)	t / U	p-value	Conclusion
1	Average sentence length (SL)	18.6 ± 3.2	21.1 ± 2.5	2.45	0.018	AI texts are longer, more structured
2	Type-Token Ratio (TTR)	0.61 ± 0.07	0.48 ± 0.05	3.90	0.001	Human texts are more lexically diverse
3	Passive structures (%)	7.8 ± 2.1	11.4 ± 2.7	2.87	0.007	AI texts use the passive voice more often
4	Emotional vocabulary (per 1000 words)	12.3 ± 3.4	5.6 ± 1.9	4.28	0.000	Human texts are much more emotional
5	Modal verbs (%)	9.1 ± 2.5	6.2 ± 1.7	2.33	0.024	AI texts demonstrate lower modality

Source: created by the author based on his own calculations using the methods Reinhart et al. (2025), Sardinha (2024), Wu (2025), and Shaib et al. (2024).

The results show statistically significant ($p < 0.05$) differences between human and AI-generated texts across five key parameters. AI texts tend to have greater grammatical consistency, such as longer sentences and more frequent passive voice, but less semantic flexibility and emotional expressiveness. Human texts display higher lexical diversity, modality, and pragmatic relevance, confirming hypothesis H_1 about systemic linguistic and stylistic differences between the two types of discourse. Graphical models based on PCA clustering (Figures 3–4) demonstrate a clear separation of the corpora, indicating that the unique human-like quality of writing remains even in highly advanced generative systems.

5. DISCUSSION

The results confirm that artificially generated media texts tend to be more grammatically correct, have standardized sentence structures, and show less use of metaphors. These findings align with Sardinha (2024) and Shaib et al. (2024), who highlight these features as key signs of machine authorship. However, our analysis somewhat differs from Curry et al. (2024), who suggest that large speech models can adapt to context when trained on multi-genre corpora. The study shows that even with such adaptability, the texts' pragmatic relevance and emotional diversity remain limited, matching the findings of Reinhart et al. (2025) and D'Andrea et al. (2025) about the models' lack of cognitive flexibility and communicative intent.

Individual authors, including Gherheș et al. (2025) and González-Arias et al. (2024), believe that the tendency toward formal neutrality in artificially generated texts is not a flaw but a reflection of a technological ethic designed to avoid emotional bias. However, our analysis shows that this neutrality reduces communicative persuasiveness in the media space, where emotional and evaluative structures are central. Lewis et al. (2025) also emphasize that journalistic credibility relies on the presence of an authorial stance, whereas algorithmic writing creates "objectivity without intention.

A certain number of researchers (Rosenfeld & Lazebnik, 2024; Rad et al., 2024) believe that the differences between human and machine texts are gradually decreasing due to improvements in context recognition algorithms. However, our study confirms the findings of Fedoriv et al. (2023) and Seals and Shalin (2023), that psycholinguistic differences-primarily in the areas of metaphorical thinking and emotional depth-remain stable even in the most recent versions of the models. These observations suggest that the cognitive complexity and multidimensionality of human discourse cannot be fully reproduced algorithmically.

It is important to note that our findings support the empirical results of Goulart et al. (2024) and Emara (2025) regarding lexical consistency and rhetorical predictability in AI-generated student and educational texts. However, unlike Sokil et al. (2022), who see the standardization of speech as a positive aspect of globalization, our data suggest that such unification reduces the text's cultural identity and weakens its communicative expressiveness. Additionally, Yanagita et al. (2024) demonstrate that pragmatic accuracy does not ensure substantive depth, which our analysis confirms: even grammatically perfect texts do not create a convincing emotional impact.

Thus, the study's results support some conclusions of earlier researchers regarding the formal correctness and structural coherence of generated texts, while also building on them. It demonstrates that human speech remains unique because of cognitive variability, culturally conditioned allusiveness, and intentional flexibility. The study also has certain limitations—it only examines media texts in English and does not consider cross-linguistic differences in stylistic structure. Future research could involve creating multilingual corpora to explore cognitive-emotional markers of "humanness" and developing ethical standards for using generative systems in journalism and media communication.

6. CONCLUSIONS AND PROSPECTS FOR FURTHER RESEARCH

The study showed that artificially created media texts are similar to human ones in terms of grammatical correctness and structural logic, but they still lack in cognitive depth, emotional expressiveness, and rhetorical variety. The novelty of this work lies in developing a detailed classification of stylistic models and linguistic markers of "humanness," which helps assess the authenticity of texts within digital media discourses. The findings improve understanding of how cognitive, pragmatic, and semantic aspects of communication interact, suggesting that the anthropomorphism of artificial text results not just from algorithmic learning but also from cultural context and information richness. Practically, this research can be used in media analysis, journalism training, and automated systems that detect machine-generated authorship. Comparing expected and actual results showed that, despite significant progress in generative models, they still cannot fully replicate the logical and emotional integration found in human discourse. A limitation of the study is its focus on a single language and the exclusion of visualmultimodal texts, which are becoming more common in today's media landscape. Future research should compare “human-likeness” across different languages, develop metrics for cognitive-semantic authenticity, and establish ethical standards for AI use in media. Ultimately, this will help create an integrated system for evaluating digital text quality, combining algorithmic precision with the depth of human language.

References

Al-Muhaissen, B. M., S. Al-Hammouri, K. M. Rachdan, and M. Habes. 2025. How AI affects the pragmatic function in media discourse: A French press perspective. *Forum for Linguistic Studies* 7(1): 369–380. https://doi.org/10.30564/fls.v7i1.7800

Batsurovska, I., N. Dotsenko, O. Gorbenko, and N. Kim. 2021. The technology of competencies acquisition by bachelors in higher education institutions in the conditions of the digital media communication environment. In *Proceedings of the International Conference on New Trends in Languages, Literature and Social Communications*, 206–213. Atlantis Press. https://doi.org/10.2991/assehr.k.210525.025

Bazaluk, O., O. Anisimov, P. Saik, V. Lozynskyi, O. Akimov, and L. Hrytsenko. 2023. Determining the safe distance for mining equipment operation when forming an internal dump in a deep open pit. *Sustainability* 15(7): 5912. https://doi.org/10.3390/su15075912 mdpi.com

Curry, N., P. Baker, and G. Brookes. 2024. Generative AI for corpus approaches to discourse studies: A critical evaluation of ChatGPT. *Applied Corpus Linguistics* 4(1): 100082. https://doi.org/10.1016/j.acorp.2023.100082

D’Andrea, A., G. Fusacchia, and A. D’Ulizia. 2025. Linguistic insights, media mechanisms and role of AI in dissemination and impact of disinformation. *Journal of Information, Communication and Ethics in Society*. https://doi.org/10.1108/JICES-01-2025-0014

Emara, I. F. 2025. A linguistic comparison between ChatGPT-generated and nonnative student-generated short story adaptations: A stylometric approach. *Smart Learning Environments* 12: 36. https://doi.org/10.1186/s40561-025-00388-z

Fedoriv, Y., I. Pirozhenko, and A. Shuhai. 2023. Linguistic analysis of human- and AI-created content in academic discourse. *Journal of Vasyl Stefanyk Precarpathian National University: Philology* 10: 47–67. https://doi.org/10.15330/jpnuphil.10.47-67

Gherheș, V., M. A. Fărcașiu, M. Cernicova-Buca, and C. Coman. 2025. AI vs. human-authored headlines: Evaluating the effectiveness, trust, and linguistic features of ChatGPT-generated clickbait and informative headlines in digital news. *Information* 16(2): 150. https://doi.org/10.3390/info16020150

González-Arias, C., E. Chatzikoumi, and X. López-García. 2024. The anthropomorphic pursuit of AI-generated journalistic texts: Limits to expressing subjectivity. *Frontiers in Communication* 9. https://doi.org/10.3389/fcomm.2024.1456509

Goulart, L., M. L. Matte, A. Mendoza, L. Alvarado, and I. Veloso. 2024. AI or student writing? Analyzing the situational and linguistic characteristics of undergraduate student writing and AI-generated assignments. *Journal of Second Language Writing* 66: 101160. https://doi.org/10.1016/j.jslw.2024.101160

Lewis, S. C., A. L. Guzman, T. R. Schmidt, and B. Lin. 2025. Generative AI and its disruptive challenge to journalism: An institutional analysis. *Communication and Change* 1(1): 9. https://doi.org/10.1007/s44382-025-00008-x

Petricini, T. 2025. The power of language: Framing AI as an assistant, collaborator, or transformative force in cultural discourse. *AI & Society*. https://doi.org/10.1007/s00146-025-02586-2

Rad, M. H., F. Farsi, S. Bali, R. Etezadi, and M. Shamsfard. 2024. RFBES at SemEval-2024 task 8: Investigating syntactic and semantic features for distinguishing AI-generated and human-written texts. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, 1–10. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.semeval-1.69

Reinhart, A., B. Markey, M. Laudenbach, K. Pantusen, R. Yurko, G. Weinberg, and D. W. Brown. 2025. Do LLMs write like humans? Variation in grammatical and rhetorical styles. *Proceedings of the National Academy of Sciences* 122(8). https://doi.org/10.1073/pnas.2422455122

Rosenfeld, A., and T. Lazebnik. 2024. Whose LLM is it anyway? Linguistic comparison and LLM attribution for GPT-3.5, GPT-4 and Bard. *arXiv:2402.14533*. https://doi.org/10.48550/arXiv.2402.14533

Sardinha, T. B. 2024. AI-generated vs human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics* 4(1): 100083. https://doi.org/10.1016/j.acorp.2023.100083

Seals, S. M., and V. L. Shalin. 2023. Long-form analogies generated by ChatGPT lack human-like psycholinguistic properties. *arXiv:2306.04537*. https://doi.org/10.48550/arXiv.2306.04537

Shah, A., P. Ranka, U. Dedhia, S. Prasad, S. Muni, and K. Bhowmick. 2023. Detecting and unmasking AI-generated texts through explainable artificial intelligence using stylistic features. *International Journal of Advanced Computer Science and Applications* 14(10). http://dx.doi.org/10.14569/IJACSA.2023.01410110

Shaib, C., Y. Elazar, J. J. Li, and B. C. Wallace. 2024. Detection and measurement of syntactic templates in generated text. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 6416–6431. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.368

Shavarskyi, I., V. Falshtynskyi, R. Dychkovskyi, O. Akimov, D. Sala, and V. Buketov. 2022. Management of the longwall face advance on the stress-strain state of rock mass. *Mining of Mineral Deposits* 16(3): 78–85. https://doi.org/10.33271/mining16.03.078 badap.agh.edu.pl

Simón, L. A., J. A. G. Gimeno, A. M. F.-P. Cesteros, M. F. Trinidad, and M. V. E. Vidal. 2023. Using linguistic knowledge for automated text identification. In *IberLEF 2023 – Proceedings of the Iberian Languages Evaluation Forum, co-located with the Conference of the Spanish Society for Natural Language Processing, SEPLN 2023*, Vol. 3496. CEUR Workshop Proceedings. https://ceur-ws.org/Vol-3496/autextification-paper17.pdf

Sokil, O., S. Kucherkova, A. Kostyakova, N. Podolchak, Y. Sokil, and N. Shkvyria. 2022. The context of "globalization versus localization" after the world pandemic and quarantine. In S. G. Yaseen (ed.), *Digital economy, business analytics, and big data analytics applications*, Vol. 1010. Springer. https://doi.org/10.1007/978-3-031-05258-3_8

Strübbe, S., I. Sidorenko, and R. Lampe. 2025. Comparison of grammar characteristics of human-written corpora and machine-generated texts using a novel rule-based parser. *Information* 16(4): 274. https://doi.org/10.3390/info16040274

Wan, Y.-N. 2024. Language differences in online complaint responses between generative artificial intelligence and hotel managers. *Informatics* 11(3): 66. https://doi.org/10.3390/informatics11030066

Wu, J. 2025. A corpus-based multidimensional analysis of linguistic features between human-authored and ChatGPT-generated compositions. *International Journal of Linguistics, Literature and Translation* 8(5): 102–110. https://doi.org/10.32996/ijllt.2025.8.5.10

Wu, J., S. Yang, R. Zhan, Y. Yuan, L. S. Chao, and D. F. Wong. 2025. A survey on LLM-generated text detection: Necessity, methods, and future directions. *Computational Linguistics* 51(1): 275–338. https://doi.org/10.1162/coli_a_00549

Yanagita, Y., D. Yokokawa, S. Uchida, Y. Li, T. Uehara, and M. Ikusaka. 2024. Can AI-generated clinical vignettes in Japanese be used medically and linguistically? *Journal of General Internal Medicine* 39(16): 3282–3289. https://doi.org/10.1007/s11606-024-09031-y

Yildiz Durak, H., F. Eğin, and A. Onan. 2025. A comparison of human-written versus AI-generated text in discussions at educational settings: Investigating features for ChatGPT, Gemini and BingAI. *European Journal of Education* 60(1). https://doi.org/10.1111/ejed.70014

Zaitsu, W., and M. Jin. 2023. Distinguishing ChatGPT-(3.5, -4)-generated and human-written papers through Japanese stylometric analysis. *PLOS ONE* 18(8). https://doi.org/10.1371/journal.pone.0288453

Zhang, M., and P. Crosthwaite. 2025. More human than human? Differences in lexis and collocation within academic essays produced by ChatGPT-3.5 and human L2 writers. *International Review of Applied Linguistics in Language Teaching*. https://doi.org/10.1515/iral-2024-0196

Zhaxylykbayeva, R., A. Burkitbayeva, B. Zhakhyp, K. Kabylgazina, and G. Ashirbekova. 2025. Artificial intelligence and journalistic ethics: A comparative analysis of AI-generated content and traditional journalism. *Journalism and Media* 6(3): 105. https://doi.org/10.3390/journalmedia6030105