

Research on Language Features of AI-generated News Texts: A Contrastive Analysis Based on Discourse Style and Syntactic Structure

Wenfang Hu

International College of Hebei University, Baoding, Hebei, 071000, China

ABSTRACT

Against the backdrop of generative AIs deep integration into news production, the linguistic characteristics of AI-generated news texts have become a key research focus. This study employs parallel corpora to conduct comparative analysis between AI-generated news and professional journalism from mainstream media outlets, utilizing corpus comparison, quantitative statistics, and discourse analysis methods to examine stylistic features and syntactic structures. Findings reveal that AI-generated news exhibits strong linguistic standardization but demonstrates stylistic homogeneity, weakened emotional expression, and pronounced lexical generalization tendencies at the stylistic level. At the syntactic level, it shows monotonous sentence patterns, uniform sentence length distribution, reliance on explicit conjunctions for logical coherence, and relatively low syntactic complexity. These characteristics reflect practical limitations in AI-generated news regarding information hierarchy construction, discursive expressiveness, and communication adaptability. The research contributes to deeper understanding of AI news language mechanisms and provides empirical references for optimizing intelligent news generation systems.

KEYWORDS

AI journalism; Stylistic features; Syntactic structure; News language; Text generation

1 Introduction

In recent years, generative artificial intelligence has been increasingly applied in news production, progressively integrating into writing, editing, and distribution processes while driving transformative changes in journalistic workflows. While extensive discussions have focused on news production mechanisms and ethical considerations, current research remains insufficient in systematically analyzing the linguistic structures of AI-generated news texts, particularly lacking comparative studies based on authentic corpora. From fundamental perspectives of news linguistics, register characteristics and syntactic structures serve as critical dimensions for revealing textual styles and information organization patterns. Existing studies predominantly concentrate on traditional news texts, showing inadequate attention to specific manifestations of language standardization, sentence structure organization, and expressive techniques in AI-generated content. Moreover, integrated research approaches combining qualitative analysis with quantitative statistics remain underdeveloped. Addressing these gaps, this study employs parallel corpus comparison as its core methodology. By examining both register characteristics and syntactic structures of AI-generated news, we systematically analyze linguistic differences from professional media reports, identify core linguistic features of AI-generated news, and explore the technical mechanisms underlying their formation alongside practical limitations.

2 Literature Review

2.1 Research on Generative Artificial Intelligence and News Production

Generative artificial intelligence (hereinafter referred to as "generative AI") is fundamentally transforming modern news production models and industrial chain structures. It is driving comprehensive changes in the core role of news producers, content creation processes, dissemination formats, and audience engagement with news feedback. This technological evolution is compelling the paradigm shift in human-machine relationships from adversarial theory to collaborative synergy^[1]. The technological evolution and current status of generative AI demonstrate that it represents not merely a tool upgrade, but a paradigm shift in production^[2].

Scholarly discussions surrounding this shift primarily focus on news authenticity and objectivity. On one hand, generative AI demonstrates strong content generation capabilities, which can be exploited to create misinformation across images, videos, and text. For instance, fabricated news content generated through deepfake technology complicates the assessment of news authenticity.[2]On the other hand, potential biases in training datasets may influence generated outputs, thereby compromising the objectivity of news content. Furthermore, studies indicate that in practical applications, some journalists might delegate decision-making responsibilities to AI tools while neglecting their own editorial oversight, which to some extent undermines the professionalism of news production^[2].

2.2 Evolutionary Journey: Integration of Generative AI and News Production

From an application perspective, generative AI has progressively deepened its role in news production. Cutting-edge models like ChatGPT demonstrate remarkable capabilities in text generation, content proofreading, information retrieval, and language translation. Research indicates that AI applications in journalism primarily focus on three key areas: efficiency enhancement, automated writing, and personalized content distribution^[3]. Traditional news production follows a linear, one-way structure where journalists and editors sequentially gather, write, edit, and distribute news to audiences. In contrast, GAI employs process mining, algorithmic resource optimization, and multi-dimensional interactive distribution to enable leapfrog connectivity across all production stages. This process collaboration ultimately transforms news production into a diversified, networked structure^[4]. Scholar Diakopoulos outlined the primary functions of GAI tools in news production practices: (1) Text processing and generation, including evaluating news value, analyzing documents for classification and tagging, handling multilingual raw data translation, reviewing and categorizing comments (e.g., "hate speech" or "high-quality comments"), creating diverse tweet styles for different social media platforms, automated writing based on unstructured/structured data, and generating/revising selected content titles/abstracts using user models. (2) Visual information editing and processing, such as editing multi-format photos/videos and generating charts based on content themes. (3) Optimizing algorithm-driven workflows, including personalized content delivery for diverse audiences and automated review/removal of inappropriate statements^[5].

2.3 Research on News Style and Language Style

Scholar Wang Yan pointed out that news discourse research begins with communicative functions, studying language features and discourse organization patterns compatible with these functions to distinguish news styles. It seeks to establish corresponding relationships between formal characteristics of news language and non-verbal environmental factors^[6]. Current research indicates that different media platforms significantly influence the stylistic characteristics of news language. Traditional media platforms such as newspapers, radio, and television exhibit distinct stylistic variations due to their unique characteristics. Newspaper news typically employs standardized, formal language, while radio news adopts a more accessible tone, and television news emphasizes seamless integration between visuals and narration. The advent of new media has revitalized news language evolution through internet-based platforms, resulting in more concise expressions, innovative vocabulary, popularized language, interactive emotional engagement, diversified sources, and personalized styles^[7]. Xing Chunhua argues that regardless of the media platform type, news language must fulfill dual missions: objectively reflecting facts while conveying emotional warmth^[8].

2.4 Research on News Syntax Structure and Language Form

Through corpus analysis, scholar Zhang Xuan pointed out that news texts exhibit certain normative characteristics in grammatical structure, often organizing information with relatively stable sentence patterns^[9]. News reporting must emphasize objectivity and authenticity. The language should reflect an objective perspective in analyzing and presenting news facts, with word choices and sentence structures strictly aligned with factual events. It must avoid exaggeration or subjective interpretations, ensuring an impartial reconstruction of news events to maintain their truthfulness. Scholar Li Lin further notes that journalistic language should embody four key characteristics: objectivity, conciseness, simplicity, and accessibility^[10].

2.5 Literature Review

Current research has explored the relationship between generative artificial intelligence and news production from multiple dimensions. At the macro level, scholars have fully recognized the profound impact of generative AI on news production paradigms, communication structures, and human-machine interactions. At the normative level, related studies focus on issues of news authenticity, objectivity, and ethical risks. At the linguistic level, research primarily examines the evolution of traditional news discourse characteristics and language styles across different media environments. However, comprehensive analysis of existing studies reveals that while existing research has conducted thorough investigations into news production mechanisms, ethical norms, and traditional news language features, systematic analysis of the ontological language structure of AI-generated news texts remains insufficient. Particularly in terms of stylistic characteristics and syntactic structures, there is a lack of comprehensive comparative corpus-based studies. To address these research gaps, this study proposes the following core questions: What are the primary linguistic expression characteristics of generative AI-generated news texts compared to traditional news writing? How does its generation mechanism influence news language forms?

3 Research Design

3.1 Research Methods

This study will conduct a comparative analysis of AI-generated news from both stylistic and syntactic perspectives, using professional news from mainstream media as a reference.

The research methodology comprises three key approaches: First, the corpus comparison method involves parallel analysis of AI-generated news articles and mainstream media reports, with controlled parameters including subject matter, genre, length, and publication timeframes to ensure validity of linguistic differences assessment. Second, quantitative statistical analysis evaluates metrics such as emotional word distribution, sentence structure patterns, average sentence length, conjunction frequency, and syntactic complexity, providing visualized evidence of language disparities between AI-generated content and traditional journalism. Third, discourse analysis examines textual stylistic features, logical coherence, and information organization through functional discourse analysis. Finally, comparative study methodology enables cross-text analysis of linguistic characteristics to identify core patterns and limitations in AI-generated news language.

3.2 Corpus Selection and Construction

To ensure the validity and comparability of the study, the corpus selection strictly adhered to the principle of "same subject matter, same genre, same length, and same time period," resulting in a parallel corpus with a total sample size of 100 articles.

AI-generated news corpus: Three leading large-scale models—ChatGPT, Wenxin Yiyang, and Tongyi Qianwen—were selected to produce 50 news articles across four categories: current affairs, social issues, livelihood topics, and finance. Guided by the unified prompt "to write objective and standardized news reports," each article was limited to 800-1000 words. Samples with generation errors or content deviations from journalistic genres were excluded to ensure textual accuracy and compliance with journalistic standards.

Mainstream media news corpus: This dataset comprises 50 news articles of identical themes and genres simultaneously published by three authoritative mainstream media outlets—Xinhua News Agency, Peoples Daily, and CCTV News. The articles match the length of AI-generated news pieces and are authored by professional journalists, representing the standard benchmarks of professional journalistic language.

3.3 Analysis Dimensions

The two core analytical dimensions are: firstly, stylistic features, including linguistic standardization, stylistic homogeneity, emotional expression tendencies, lexical selection characteristics, and textual formality; secondly, syntactic structures, encompassing sentence type distribution, sentence length distribution, logical cohesion patterns, syntactic complexity, and information hierarchy construction.

4 Comparison of Discourse Features between AI-generated News and Professional News

4.1 AI-generated News Exhibits Highly Standardized Language but Homogeneous Style

AI-generated news demonstrates exceptional linguistic standardization, with texts free from grammatical errors, lexical inaccuracies, or logical inconsistencies. These outputs fully comply with modern Chinese written language norms and journalistic stylistic requirements, achieved through deep learning of massive standardized texts by large-scale AI models that enable error-free text generation even for complete language beginners. However, this comes with a notable drawback: news writing exhibits highly homogeneous stylistic patterns. AI-generated content consistently adopts neutral expression styles across all topics and communication scenarios, showing no discernible differences in vocabulary selection, tone modulation, or discourse intonation. This lack of contextual adaptation for varying news themes and communication environments contrasts sharply with professional journalism. In contrast, mainstream media journalists produce politically charged news with solemn rigor, community-focused reports with accessible language, and financial news characterized by professional conciseness. Their writing styles demonstrate precise alignment with subject matter and target audiences, exemplifying the situational adaptability and targeted precision inherent in professional journalism.

4.2 AI-generated News Exhibits Weakened Emotional Expression and Lacks Humanistic Warmth

The neutrality and objectivity inherent in journalistic discourse do not entail the complete exclusion of human emotions. Rather, they involve restrained and measured value judgments alongside humanistic considerations grounded in factual events—this constitutes the core characteristic of professional news communication. AI-generated news mechanically equates journalistic neutrality with "emotion removal." Comparative analysis reveals that such texts contain minimal emotional language, biased terminology, or subjective evaluations, maintaining an entirely fact-based narrative structure. This results in pronounced emotional detachment and a lack of humanistic warmth in their expression.

While AI-generated news achieves extreme objectivity by avoiding subjective biases, it also strips texts of the emotional warmth and persuasive power inherent to journalism. This undermines the dual mission of professional news—delivering facts while guiding values. Consequently, such content struggles to resonate emotionally with audiences in public discourse or civic news narratives, failing to cultivate empathy.

4.3 There is a Pronounced Preference for Lexical Generalization, with Insufficient Scenario-specific Precision in Expression

In terms of vocabulary selection, AI-generated news heavily relies on generic journalistic terms, favoring neutral words with high frequency and broad applicability while lacking professional industry-specific terminology. The overall lexicon tends to adopt "safety-oriented and generalized" expressions. Although free from lexical errors, it falls short in precision and contextual relevance. In contrast, professional media news emphasizes lexical accuracy and situational appropriateness, employing specialized terminology tailored to event attributes and incorporating concrete, detail-rich expressions to meet narrative demands. From the audiences perspective, journalistic language demonstrates stronger visual appeal, aligns with on-site storytelling requirements, and exhibits greater professional journalistic sophistication in vocabulary selection.

5 Syntactic Structure Comparison Between AI-generated News and Professional News

5.1 AI-generated News Exhibits Monotonous Sentence Structures with Limited Syntactic Rhythm Variation

Syntax and sentence structures serve as crucial vehicles for news information organization. Professional journalism employs strategic sentence variations to effectively prioritize information and control narrative pacing. In contrast, AI-generated news exhibits highly uniform sentence patterns, with declarative sentences dominating significantly over professional journalism. It scarcely utilizes rhetorical devices like rhetorical questions, inverted structures, or emphatic expressions, relying predominantly on simple sentences and basic coordinate constructions while employing compound sentences and complex structures with minimal frequency. The resulting text maintains rigid syntactic rhythms devoid of dynamic variation. Mainstream media news, however, builds upon declarative foundations by skillfully incorporating compound sentences, short phrases, and specialized sentence patterns. This interplay of varied sentence lengths and sentence types ensures clear information delivery while emphasizing core messages through rhythmic syntactic variations, thereby enhancing the narrative depth of news reporting.

5.2 AI-generated News Sentences Exhibit Uniform Sentence Length Distribution Without Hierarchical Information Rhythm Design

Quantitative analysis reveals that AI-generated news exhibits concentrated average sentence lengths with highly uniform text structures, demonstrating algorithmic characteristics of textual homogeneity. In contrast, professional journalism demonstrates flexible sentence structures: concise sentences capture core information while extended paragraphs elaborate background details, with alternating sentence lengths better aligning with readers reading habits. Hierarchical information differentiation is achieved through variations in sentence length distribution. Although the uniform sentence length in AI-generated news maintains textual fluency, it fails to distinguish between primary and secondary information through sentence length variations, making it challenging for audiences to quickly identify key news points.

5.3 The logical coherence in AI-generated news relies on explicit conjunctions, resulting in rigid and mechanical transitional patterns

In terms of discourse logical cohesion, AI-generated news heavily relies on explicit conjunctions such as "firstly," "secondly," "therefore," and "simultaneously." The frequency of conjunction usage is significantly higher than in professional journalism, with sentences and paragraphs achieving logical connections solely through external markers. This approach lacks natural transitions, resulting in forced articulation and superficial linguistic coherence. Professional journalism prioritizes semantic internal connections, employing more precise conjunctions and relying on implicit cohesion through factual logic and narrative sequencing. Consequently, transitions within discourse flow more naturally and smoothly.

6 Generative Mechanism and Practical Limitations of AI News Language Features

6.1 Intrinsic Generation Mechanism

The probabilistic prediction mechanism of artificial intelligence algorithms dictates that AI tends to select the most universal, compliant, and statistically probable language expressions to ensure textual fluency and standardization. This

results in AI-generated news content exhibiting standardized, homogeneous, and monotonous sentence structures. Secondly, model training data predominantly consists of standardized written texts, lacking practical journalistic knowledge such as news context adaptation, stylistic adjustments, and narrative techniques. Consequently, these models fail to acquire the flexibility and targeted precision required for professional news writing.

6.2 Practical Limitations

Firstly, the lack of information hierarchy construction capability results in monotonous sentence structures, rigid sentence lengths, and forced transitions. This prevents readers from distinguishing between primary and secondary news information through linguistic frameworks, leading to diluted core messages that fail to meet audiences demand for quick news consumption and essential content retrieval. Secondly, insufficient narrative expressiveness manifests in homogeneous styles, emotional detachment, and generalized vocabulary, lacking the communicative tension and emotional resonance characteristic of professional journalism, thereby hindering effective news dissemination. Thirdly, blurred professional discourse identity persists as AI-generated news remains confined to factual organization and standardized expression, devoid of journalists critical judgment, narrative intent, and disciplinary perspectives, failing to establish authentic journalistic discourse systems. Fourthly, inadequate contextual adaptability prevents AI news from adjusting language styles according to media platforms and audience demographics, making it ill-suited for diverse news communication scenarios.

7 Conclusion

This study conducts a comparative analysis of stylistic features and syntactic structures between AI-generated news and mainstream media reports based on parallel corpora. The findings reveal that AI-generated news demonstrates stable linguistic standardization, yet exhibits characteristics such as stylistic homogeneity, limited emotional expression, and generalized vocabulary usage. At the syntactic level, it predominantly employs simple sentences with restricted sentence pattern variations, balanced sentence length distribution, and text cohesion relying mainly on explicit linking devices. In contrast, mainstream media news demonstrates greater flexibility in stylistic application and syntactic organization, enabling effective differentiation of information priorities and hierarchical expression through linguistic form modulation.

From a generative perspective, these differences are closely tied to the current language generation methods of generative models and their training corpus structures. Overall, AI-generated news has achieved basic standardized expression capabilities, but there remains room for improvement in information organization and expression levels, requiring continuous optimization through practical applications.

References

- [1] Liu Yongmou, Wang Chunli. Human-Machine Relationships in the Intelligent Era: A Theory of Choice Toward Technological Control [J]. *Global Media Journal*, 2023,10(3):5-21.
- [2] Li Xiaomeng. Ethical Dilemmas and Governance Framework of Generative AI in the Field of News Communication [J]. *Radio and Television Network*, 2026,33(03):61-63.
- [3] Monojoy Bhattacharjee. Almost 50% of news publishers use Generative AI tools, but "quality of content" is the #1 concern [EB/OL]. 2023-01-23 [2023-07-14]. <https://whatsnewinpublishing.com/almost-50-of-news-publishers-use-generative-ai-tools-but-quality-of-content-is-the-1-concern/>.
- [4] Jiang Xueying, Liu Xin, Xu Jing. Frontier Applications and Regulatory Approaches of Generative AI News from a Human-Machine Collaboration Perspective [J]. *News Enthusiasts*, 2023, (11):38-43.
- [5] Nick Diakopoulos. What Could ChatGPT Do for News Production? [EB/OL]. 2023-01-15 [2023-07-15]. <https://generative-ainewsroom.com/what-could-chatgpt-do-for-news-production-2b2a3b7047d9>.
- [6] Wang Yan. Review of Research on News Language [J]. *Rhetoric Learning*, 2002(6):28-29.
- [7] Huang Shanwei. Characteristics of News Language and Its Evolution in the New Media Era [J]. *Journalist Cradle*, 2024, (06):18-20.
- [8] Xing Chunhua. Teaching Innovation Based on the "Modern Chinese" Course for Journalism and Communication Majors [J]. *China Journal of Multimedia and Network Teaching (First Ten-Day Issue)*, 2025, (10):193-196.
- [9] Zhang Xuan. Quantitative Study on the Syntactic Structure Features of Chinese English in News Reports [D]. Guangxi Normal University, 2004.
- [10] Deng Jun. Intertextual Relationships in Interpretive Journalism: A Linguistic Perspective on the Concept of Intertextuality [J]. *Contemporary Rhetoric*, 2011(05):42-55.