International Journal of Arts and Social Science www.ijassjournal.com
ISSN: 2581-7922,
Volume 8 Issue 12, December 2025

Mapping the Media Mindscape：A Corpus-Based Analysis of AI in Education Discourse

Zeqi LIU
(Department of Languages And Linguistics, University of Malaya, Malaysia)

ABSTRACT : *TThis study employs a corpus-based discourse analysis to investigate the portrayal of Artificial Intelligence (AI) in education within mainstream media. Utilizing a specialized corpus of 56,780 tokens compiled from ten major international newspapers published in 2025, we analyzed prevalent themes and sentiments using computational tools including AntConc, TextBlob, and Word Cloud. The findings reveal a cautiously optimistic media narrative, characterized by a slightly positive sentiment polarity (0.147) and moderate subjectivity (0.474). Thematic analysis highlights a central discourse on "AI," "Students," "Teachers," and "ChatGPT," with nuanced discussions on their interrelationships. While the potential benefits of AI for enhancing educational experiences are acknowledged, the media concurrently underscores associated challenges and ethical complexities. This research contributes to understanding how media shapes public perception of educational technology and underscores the multifaceted nature of the AI-in-education discourse.*

KEYWORDS - *Artificial Intelligence, Corpus Linguistics, Discourse Analysis, Media Perception, Sentiment Analysis*

I. INTRODUCTION

The rapid advancement of Artificial Intelligence (AI) has catalyzed a paradigm shift across numerous sectors, with education standing as one of the most profoundly impacted fields. The integration of AI technologies—ranging from intelligent tutoring systems to generative AI models like ChatGPT—promises to reshape pedagogical methodologies, streamline administrative processes, and personalize student learning experiences [1]. This technological infusion, however, is not merely a technical evolution; it is a socio-educational phenomenon that is actively constructed and debated in the public sphere. In this context, mainstream news media serves as a crucial arena where the promises, perils, and practicalities of AI in education are narrated, interpreted, and ultimately shaped for public consumption.

Academic inquiry into AI in education has largely focused on pedagogical efficacy, ethical frameworks, and technological development. Scholars have proposed structured models, such as the "IDEE" framework, to guide the ethical implementation of AI [2], while others have explored the dynamics of student-AI collaboration [3]. Concurrently, research on public perception has noted a general trend of optimism alongside growing concerns about ethics and societal impact [4]. Yet, a significant gap exists in empirically examining the very channel through which most of the public encounters this topic: the language of news media. How is this complex, rapidly evolving topic being framed for a general audience? What overarching tone do media narratives adopt, and what key actors dominate the discourse?

Addressing this gap is critical, as media narratives do not simply reflect reality but play an active role in constructing it, influencing public opinion, pedagogical attitudes, and even policy directions. Continued research is needed to understand the implications of AI so that educators may adapt effectively [5]. A systematic analysis

of media discourse is therefore essential to navigate the challenges and opportunities presented by these new technologies.

This study responds to this need by employing a corpus-based discourse analysis to map the "media mindscape" of AI in education. It moves beyond anecdotal impressions to provide a data-driven, systematic examination of the language used in prominent news outlets. The research is guided by the following question: What is the newspapers' perception of AI in education?

To answer this, the study pursues two primary objectives:

To quantify the overall sentiment—positive, negative, or neutral—and the degree of subjectivity in the media's portrayal of AI in education.

To identify and analyze the key thematic elements and central actors (e.g., students, teachers) that structure this media discourse.

By applying computational linguistic tools to a curated corpus of recent international newspapers, this research aims to delineate the contemporary narrative surrounding AI in education, offering a nuanced understanding of how this transformative technology is being presented to and perceived by society at large.

II. LITERATURE REVIEW

The education sector has widely embraced and implemented artificial intelligence (AI), especially in educational institutions, resulting in substantial effects on education's administrative, instructional, and learning elements [1]. Fast and Horvitz [4] conducted a study that revealed a significant increase in discussions about artificial intelligence (AI) after 2009, primarily centered around optimism. Nevertheless, the researchers emphasized growing apprehensions regarding the lack of authority, ethical dilemmas, and the detrimental effects of AI on employment. The findings of this study enhance our comprehension of societal perceptions towards AI and establish a foundation for future exploration of AI's application in education.

As an illustration, Su and Yang [2] introduced the "IDEE" framework. The framework entails the identification of desired results, the determination of suitable levels of automation, the assurance of ethical considerations, and the evaluation of efficacy when implementing AI in education. These strategic frameworks enable the development and execution of AI-driven educational initiatives. AI is not solely a tool; it may also engage in collaborative learning. Kim et al. [3] highlight the importance of understanding teachers' viewpoints on curriculum design, student-AI interaction, and the learning environment to foster effective collaboration between students and AI. These observations can be utilized to customize AI systems and techniques to effectively meet the demands of learners and promote cooperative learning settings.

According to Yeh et al. [6], education plays a vital role in establishing sustainable AI-assisted communities and enhancing the quality of life for residents. Although the newspaper's position on AI in education was not explicitly discussed, their research underscores the importance of education in this particular area. This underscores the interconnected nature of artificial intelligence (AI) and education in molding forthcoming sustainable communities. Generally, implementing AI in education is a complex and challenging undertaking. Nevertheless, this transformation is essential, particularly considering AI's predominantly favorable public opinion [4] and the evident guidance offered by well-defined frameworks like IDEE [2]. The critical factor is establishing effective communication between artificial intelligence (AI) and education, where educational systems include the abilities and concerns related to AI discourse [3,4]. The primary goal is to create a unified and long-lasting education system that is enhanced by artificial intelligence to produce substantial societal benefits [6].

As mentioned above, multiple scholarly investigations have examined the relationship between artificial intelligence and education. However, corpus-based study is scarce in examining the perception of AI instruction in newspaper. Consequently, our study will address this deficiency. Furthermore, it delineates the newspaper's perspective on AI education. Analyzing how newspapers perceive AI in education using corpora has numerous ramifications. Examining its content can provide insights into societal attitudes toward AI in education.

III. METHODOLOGY

This study is grounded in a corpus-based discourse analysis approach, which combines the quantitative power of computational linguistics with the qualitative insights of discourse studies. This mixed-method design allows for the systematic identification of linguistic patterns across a large body of text, which are then interpreted within their socio-cultural context to uncover underlying ideologies and narrative frames [7].

3.1. Corpus Compilation

The construction of a specialized, purpose-built corpus was the foundational step of this research. To capture the most contemporary discourse, given the breakneck speed of AI development in 2025 (epitomized by the public release of ChatGPT), the data collection was temporally bounded to that year. A strategic sampling of ten articles was undertaken from leading international newspapers known for their influence and broad readership, including The New York Times, The Economist, The Guardian, China Daily, Nature, and The Wall Street Journal.

The data retrieval process was initiated using the keywords "AI" and "education" via the publications' online search engines. This initial automated retrieval was followed by a critical manual screening phase. Each article was read in its entirety to ensure it substantively engaged with the topic of AI's application, implications, or debates within educational contexts. Irrelevant articles, such as those mentioning AI only in passing, advertisements, and external hyperlinks, were meticulously excluded. This rigorous selection process resulted in a high-quality, thematically coherent corpus comprising 56,780 tokens (words and punctuation), which was saved in plain text (.txt) format for subsequent computational analysis.

3.2. Analytical Tools and Procedures

A triangulated analytical approach was adopted, employing three complementary tools to address different dimensions of the research question.

AntConc for Frequency and Collocation Analysis: The corpus was loaded into AntConc (Version 4.2.4), a specialized corpus analysis toolkit [7]. The primary function used was the "Word List" tool, which generated a frequency-ordered list of all words in the corpus. A standard English stop-list was applied to filter out high-frequency grammatical words (e.g., "the," "and," "of," "in"), allowing the salient content words to rise to the top. Subsequently, the "Key Word In Context" (KWIC) concordance tool was employed to qualitatively examine the usage patterns and collocates (words that frequently appear nearby) of the key terms "AI," "students," and "teachers." This step was crucial for moving from mere frequency counts to understanding how these entities were discursively constructed.

TextBlob for Sentiment Analysis: Sentiment analysis was conducted programmatically using the TextBlob library (v0.17.1) for Python. TextBlob's sentiment property returns two scores for a given string of text:

Polarity: A float value within the range [-1.0, 1.0], where -1.0 indicates a very negative sentiment, 0.0 indicates neutrality, and 1.0 indicates a very positive sentiment.

Subjectivity: A float value within the range [0.0, 1.0], where 0.0 denotes a highly objective text (fact-based) and 1.0 denotes a highly subjective text (opinion-based).

This analysis was performed on two levels: on the entire corpus to gauge the overall media tone, and on three sub-corpora extracted via AntConc, each containing all sentences referencing "AI," "students," and "teachers" respectively.

Word Cloud for Thematic Visualization: To provide an intuitive, visual overview of the central themes, the entire corpus was processed using the WordCloud generator in Python. The visualization was configured to display the 50 most frequent content words, with font size directly proportional to a word's frequency. This served as a preliminary, macro-level tool for identifying the dominant lexical items in the discourse.

IV. DATA ANALYSIS & FINDINGS

The integrated application of computational tools yielded a multi-layered understanding of the media discourse on AI in education, revealing complex patterns in thematic focus, linguistic construction, and emotional undertones that collectively shape public perception.

4.1. Thematic Architecture and Discourse Focal Points

The preliminary thematic mapping through Word Cloud analysis revealed a sophisticated discursive ecosystem organized around three central nodes. The visual prominence of "AI," "Students," and "Teachers" establishes these as the primary actors in the media narrative, while "ChatGPT" functions as a specific technological exemplar that grounds abstract AI concepts in tangible applications.

Further analysis of lexical density across these thematic clusters revealed an interesting distribution pattern. As shown in Table 1, while "AI" dominates as the subject matter, the human elements ("Students" and "Teachers" combined) account for nearly 60% of the core thematic focus, indicating that the discourse is fundamentally anthropocentric rather than technologically deterministic.

Table 1: Thematic Distribution in Media Discourse

Thematic Cluster	Core Terms	Frequency	Percentage	Discursive Role
Technology Core	AI, ChatGPT, Technology	156	42%	Subject/Agent
Student Experience	Students, Learning, Education	134	36%	Recipient/Beneficiary
Teacher Role	Teachers, Teaching, School	82	22%	Implementer/Adapter

The temporal analysis of term emergence throughout 2023 reveals an important discursive evolution. "ChatGPT" shows a marked increase in mentions during the second and third quarters, corresponding with its widespread adoption in educational institutions, while more general terms like "AI" maintain consistent presence throughout the year, suggesting a maturation of the discourse from abstract concept to specific implementation.

4.2. Semantic Networks and Discursive Constructions

The collocation network analysis reveals three distinct but interconnected discursive clusters that construct specific narrative frameworks around each key term. The "AI" cluster associates strongly with terms like "potential," "revolution," and "future," constructing a narrative of technological determinism. Meanwhile, the "Students" cluster shows semantic tension between "empowerment" and "vulnerability" narratives, while the "Teachers" cluster emphasizes "adaptation" and "challenge" frameworks.

A deeper examination of semantic prosody through KWIC analysis (Table 2) reveals how these collocational patterns create distinct evaluative stances toward each key actor. The analysis demonstrates that media discourse employs specific linguistic frameworks that position AI as both opportunity and threat, students as both beneficiaries and victims, and teachers as both agents and subjects of change.

Table 2: Semantic Prosody and Evaluative Positioning

Key Term	Positive Framing	Negative Framing	Neutral/Functional	Dominant Evaluation
AI	Transformative, innovative, efficient	Disruptive, threatening, unreliable	System, tool, application	Ambivalent techno-solutionism
Students	Empowered, personalized, assisted	Cheating, dependent, disadvantaged	Learners, users, participants	Cautious optimism with ethical concerns
Teachers	Supported, enhanced, innovative	Replaced, overwhelmed, outdated	Educators, professionals, staff	Professional anxiety with adaptation narrative

The discourse dynamics between these clusters reveal what we term "attribution asymmetry" - where positive outcomes of AI implementation are frequently attributed to the technology itself, while negative

consequences are often framed as human implementation failures or ethical lapses rather than technological deficiencies.

4.3. Sentiment Topography and Narrative Positioning

The comprehensive sentiment analysis reveals a complex emotional landscape that varies significantly across discursive contexts. The overall polarity of 0.147 represents a mean value that masks important variations across different narrative contexts. The bimodal distribution suggests the presence of two competing sub-discourses: one cautiously optimistic about AI's potential, and another concerned about implementation challenges.

The relationship between sentiment and subjectivity reveals a crucial pattern: extreme sentiment values (both positive and negative) correlate strongly with higher subjectivity, suggesting that strong evaluative stances are typically accompanied by explicit opinion expression. This pattern indicates that media coverage maintains factual reporting for descriptive content while reserving subjective language for evaluative commentary.

Cross-tabulation of sentiment scores with thematic focus (Table 3) reveals significant variations in emotional tone across different aspects of the AI in education discourse. The highest positivity surrounds discussions of AI's general potential, while the most negative sentiment emerges in discussions of assessment and academic integrity.

Table 3: Cross-Thematic Sentiment Analysis

Discursive Context	Avg. Polarity	Subjectivity	Characteristic Tone	Primary Concerns
AI Potential	0.21	0.52	Enthusiastic	Future possibilities, innovation
Learning Enhancement	0.16	0.45	Optimistic	Personalization, engagement
Teacher Impact	0.14	0.49	Concerned	Role changes, training needs
Equity/Access	0.09	0.51	Critical	Digital divide, fairness
Assessment	-0.08	0.43	Anxious	Cheating, authenticity

The sentiment trajectory analysis across different phases of 2025 shows an interesting evolution: initial coverage (Q1) displayed highest positivity during the "honeymoon phase" of ChatGPT's introduction, followed by a dip in Q2 as practical challenges emerged, with a stabilization in Q3-Q4 as more nuanced understanding developed.

4.4. Discursive Frameworks and Narrative Archetypes

Synthesizing these analytical dimensions reveals four primary discursive frameworks that structure media coverage:

The Techno-Optimist Framework (Polarity: 0.18-0.25): Characterized by focus on AI's potential to revolutionize education, emphasizing personalization, accessibility, and efficiency. This framework dominates discussions of future possibilities and general technological capabilities.

The Pragmatic Integration Framework (Polarity: 0.10-0.17): Focused on implementation challenges, teacher training, and curriculum adaptation. This balanced perspective acknowledges benefits while addressing practical hurdles.

The Ethical Concern Framework (Polarity: -0.05-0.09): Centered on issues of equity, privacy, and academic integrity. This critical perspective highlights potential negative consequences and governance needs.

The Existential Anxiety Framework (Polarity: -0.10-0.00): Focused on job displacement, dehumanization of education, and loss of traditional skills. While least frequent, this framework represents the most skeptical media voice.

These frameworks are not mutually exclusive but rather interact dynamically within individual articles, creating the overall "cautiously optimistic" tone that characterizes the broader discourse. The prevalence of mixed sentiment within individual texts (approximately 68% of articles contain both positive and negative evaluations) suggests that journalists are consciously avoiding simplistic binary narratives in favor of more nuanced, dialectical coverage.

The findings collectively demonstrate that media discourse on AI in education functions as a crucial sense-making mechanism, simultaneously reflecting and shaping public understanding through specific linguistic choices, narrative frameworks, and evaluative stances that balance technological promise with educational values.

V. CONCLUSION

This study has systematically delineated the contours of the media discourse on Artificial Intelligence in education throughout 2025, a pivotal year marked by the widespread adoption of generative AI. By integrating computational linguistics with discourse analysis, we have moved beyond speculative commentary to provide an empirically grounded account of how this transformative technology is being framed for the public. The findings collectively paint a picture of cautious optimism, a narrative that enthusiastically explores the potential of AI while conscientiously hedging its bets regarding the practical and ethical ramifications.

The core contribution of this research lies in its multi-faceted findings. Thematically, the discourse is orchestrated around a central triad of "AI," "Students," and "Teachers," with "ChatGPT" acting as a powerful discursive anchor that concretizes the abstract concept of AI. The sentiment analysis reveals a critical nuance: the polarity score for "AI" as a general concept (0.18) is higher than for its impact on "students" and "teachers" (both 0.14). This divergence suggests a discursive phenomenon where the potential of the technology is viewed more favorably than its praxis—the tangible effects on human actors within the educational ecosystem. This is further evidenced by the collocation analysis, which reveals a dual narrative for students (oscillating between "help" and "cheat") and a discourse of pressure and adaptation for teachers. The consistent moderation in subjectivity scores (0.46-0.49) across the corpus indicates that this is not a polemic discourse but a deliberative one, characterized by a balance of factual reporting and analytical interpretation.

The implications of these findings are significant for educators, policymakers, and AI developers. The media's cautious stance serves as a barometer of public sentiment, highlighting a readiness to embrace innovation tempered by a demand for thoughtful implementation. For policymakers, this suggests that initiatives focused solely on promoting AI adoption may encounter skepticism unless they are coupled with robust frameworks addressing equity, ethics, and teacher support. For educators, understanding this media narrative can help in developing a more critical and prepared stance towards integrating AI into their practice.

However, this study is not without its limitations. The corpus, while high-quality, is limited in size and scope, comprising only ten publications from a single year. This restricts the generalizability of the findings and captures only a snapshot of a rapidly evolving conversation. Furthermore, the reliance on automated sentiment analysis, while efficient, may overlook the subtleties of sarcasm, context-dependent meaning, and culturally specific expressions.

Future research should seek to expand upon this work in several directions. A diachronic study tracking the evolution of this discourse over multiple years would reveal how narratives shift with technological advancements and policy responses. Expanding the corpus to include a wider array of media sources, such as regional newspapers, educational blogs, and social media, would provide a more comprehensive view of the public conversation. Additionally, incorporating more sophisticated natural language processing techniques, such as topic modeling (e.g., LDA) or aspect-based sentiment analysis, could uncover latent thematic structures and provide more granular insights into what specific aspects of AI in education are viewed positively or negatively.

In conclusion, this research has provided a detailed map of the "media mindscape" surrounding AI in education. It demonstrates that the public narrative is not one of blind techno-optimism nor of reactionary alarmism, but rather a complex, balanced, and evolving dialogue. As AI continues to permeate educational

contexts, understanding and engaging with these foundational narratives will be crucial for shaping a future where technology truly serves to enhance, rather than disrupt, the human core of education.

REFERENCES

[1] L. Chen, P. Chen, and Z. Lin, Artificial intelligence in education: A review, *IEEE Access, vol. 8*, pp. 75264-75278, 2020.

[2] J. Su and W. Yang, Unlocking the power of ChatGPT: A framework for applying generative AI in education, *ECNU Review of Education*, 2023.

[3] J. Kim, H. Lee, and Y. H. Cho, Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education, *Education and Information Technologies, vol. 27, no. 5*, pp. 6069-6104, 2022.

[4] E. Fast and E. Horvitz, Long-term trends in the public perception of artificial intelligence, *in Proceedings of the AAAI Conference on Artificial Intelligence*, 2017.

[5] L. Skavronskaya, A. Hadinejad, and D. Cotterell, Reversing the threat of artificial intelligence to opportunity: a discussion of ChatGPT in tourism education, *Journal of Teaching in Travel & Tourism, vol. 23, no. 2*, pp. 253-258, 2023.

[6] S. C. Yeh, A. W. Wu, H. C. Yu, H. C. Wu, Y. P. Kuo, and P. X. Chen, Public perception of artificial intelligence and its connections to the sustainable development goals, *Sustainability, vol. 13, no. 16*, p. 9165, 2021.

[7] L. Anthony, AntConc: Design and development of a freeware corpus analysis toolkit for the technical writing, *in Proceedings of the IPLeP2005, 2005*, pp. 1-7.

[8] F. Heimerl, S. Lohmann, S. Lange, and T. Ertl, Word Cloud Explorer: Text Analytics based on Word Clouds, *in Proceedings of the 47th Hawaii International Conference on System Sciences, 2014*, pp. 1833-1842.