The Em Dash —A Small Mark of a Bigger Shift

ARTICLE

The Em Dash as a Site of Contest Between AI Determinism and Human Agency

This study examines the phenomenon of increased em dash usage in AI-generated text and its subsequent influence on human writing practices through the lens of technological determinism and social construction of technology (SCOT).

Eric Hawkinson
Learning Futurist
erichawkinson.com

The Em Dash—A Small Mark of a Bigger Shift

"We shape our tools, and thereafter our tools shape us."

This Marshall McLuhan insight frames an essential discussion I lead in my graduate course on educational technology in distance education—specifically the tension between technological determinism and the social construction of technology (SCOT).

To help make these concepts relatable, I often share a personal example: the em dash. I used to overuse ellipses... not always correctly. But over time, I noticed something—AI-generated text often leans heavily on em dashes. After reading and working with language models regularly, I found myself doing the same.

Two Ways of Looking at Technology

Technological determinism suggests that technology drives social change—that innovations emerge and inevitably reshape how we think, communicate, and learn. The em dash shift in my writing would be seen as proof that AI tools are fundamentally altering human expression, regardless of our intentions.

The social construction of technology argues the opposite: that human needs, desires, and social forces shape how technologies develop and get used. From this perspective, my adoption of em dashes reflects deliberate choices about efficiency and style, not technological control.

But as I teach my students, these aren't mutually exclusive forces—they're part of an ongoing relationship. We create technologies that then change us, leading us to create new technologies based on our transformed state. My em dash evolution illustrates this perfectly.

Connection to the Automation Abyss

This seemingly trivial punctuation shift illustrates something much larger: the Automation Abyss opening between those who direct AI and those directed by it. When I unconsciously adopt em dashes, I'm being directed by algorithmic patterns embedded in my tools. When I recognize this pattern and analyze it critically—as I'm doing now—I'm maintaining agency over the technology.

The difference matters enormously. Students who blindly adopt AI-generated writing patterns without understanding them risk losing their own voice. Those who recognize these influences and make deliberate choices about when to embrace or resist them maintain their intellectual autonomy. The same dynamic plays out whether we're talking about punctuation marks, research methods, or fundamental learning processes.

In education, we're witnessing this relationship scale from punctuation to pedagogy. Students increasingly rely on AI not just for writing assistance but for brainstorming, analysis, and even basic inquiry. Each convenience comes with a trade-off: efficiency gained, capability potentially lost. The challenge isn't to reject these tools but to design learning experiences that preserve human agency within increasingly automated environments.

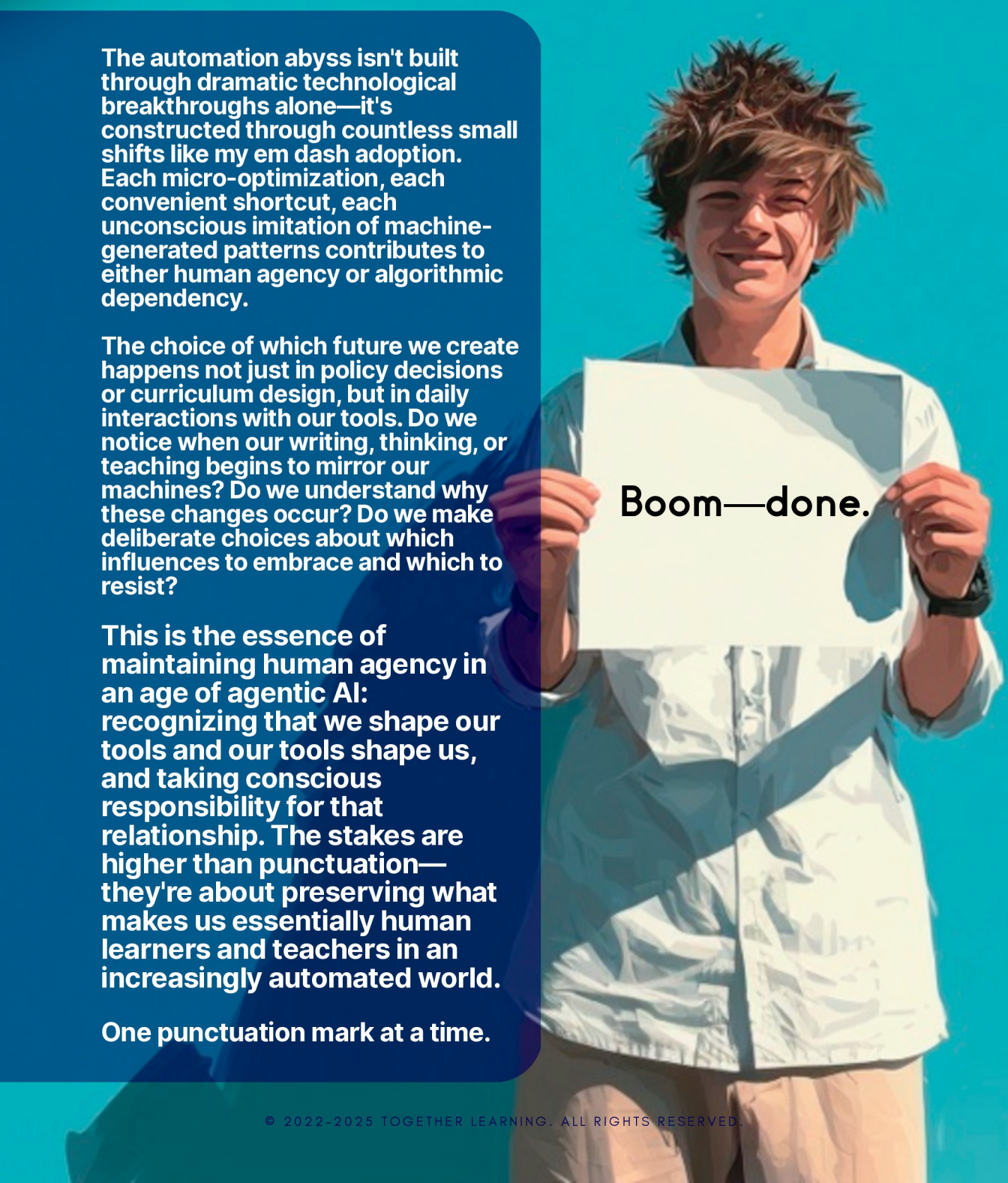
One Punctuation Mark at a Time

The automation abyss isn't built through dramatic technological breakthroughs alone—it's constructed through countless small shifts like my em dash adoption. Each micro-optimization, each convenient shortcut, each unconscious imitation of machine-generated patterns contributes to either human agency or algorithmic dependency.

The choice of which future we create happens not just in policy decisions or curriculum design, but in daily interactions with our tools. Do we notice when our writing, thinking, or teaching begins to mirror our machines? Do we understand why these changes occur? Do we make deliberate choices about which influences to embrace and which to resist?

This is the essence of maintaining human agency in an age of agentic AI: recognizing that we shape our tools and our tools shape us, and taking conscious responsibility for that relationship. The stakes are higher than punctuation—they're about preserving what makes us essentially human learners and teachers in an increasingly automated world.

One punctuation mark at a time.



ABOUT THE AUTHOR

ERIC HAWKINSON

Eric is a learning futurist, tinkering with and designing technologies that may better inform the future of teaching and learning. Eric's projects have included augmented tourism rallies, AR community art exhibitions, mixed reality escape rooms, and other experiments in immersive technology.

Eric Hawkinson
LEARNING FUTURIST

erichawkinson.com

Roles

Professor – Kyoto University of Foreign Studies
Research Coordinator – MAVR Research Group
Founder – Together Learning
Developer – Reality Labo
Community Leader – Team Teachers
Chair – World Learning Labs

CORE VALUES

- Open Knowledge - Free and open access to information is a foundation to a productive modern life, connected to ideas of the open web and platform agnosticism.
- Privacy by Design - Business models are increasing moving toward supporting revenue by collecting, curating, and trading behavioral surplus through technology. These models should be tempered with safety, ethics, and privacy concerns and designed as such.
- Digital Literacy for All - An informed public about the use of technology is key for a responsible and engaged digital society.

PASSION PROJCTS

Together Learning
A community of technology minded learners. Exploring human potential... together.

Reality Labo
Augmented reality enhanced learning environments and mixed reality rapid prototyping.

Corefol.io
Showcase the Core of your Skills with AI Assisted Student Portfolios

QUESTION TO YOU

What/how do you determine as ‘too much ai’ when thinking about assessing a student writing submissions?

The Em Dash as a Site of Contest Between AI Determinism and Human Agency

Eric Hawkinson (Professor and Learning Futurist)
Institute of Liberal Arts and Studies, Kyoto University of Foreign Studies
Kyoto, Japan
ORCID: 0000-0003-4187-6992 – https://erichawkinson.com

***Abstract*— This study examines the phenomenon of increased em dash usage in AI-generated text and its subsequent influence on human writing practices through the lens of technological determinism and social construction of technology (SCOT). Through analysis of recent empirical research, this paper addresses two primary research questions: (1) What computational and data-driven factors explain the prevalence of em dashes in major LLM outputs? and (2) How are educational institutions and writers responding to this phenomenon? The research reveals a complex dialectical relationship where computational factors including training data composition, tokenization processes, and reinforcement learning from human feedback—create distinctive punctuation patterns that subsequently influence human writing behavior. Educational institutions are adapting through comprehensive policy frameworks while writers are modifying their practices in response to AI-generated text characteristics. The findings support Marshall McLuhan's concept that "we shape our tools, and thereafter our tools shape us,"[11] demonstrating that AI writing tools represent neither pure technological determinism nor complete social construction, but rather an ongoing co-evolutionary process reshaping the fundamental nature of written communication.**

Keywords*— *artificial intelligence, writing pedagogy, punctuation, technological determinism, social construction of technology, large language models

I. Introduction

The emergence of sophisticated AI writing tools has catalyzed a fundamental transformation in how humans create and interact with written text. Among the most intriguing manifestations of this shift is the widespread observation that AI-generated text exhibits distinctive punctuation patterns, particularly an increased frequency of em dashes that has become so pronounced it has earned the colloquial designation "ChatGPT hyphen" in popular discourse. This phenomenon represents more than a mere stylistic curiosity; it embodies a complex dialectical relationship between technological capabilities and human writing practices that demands rigorous academic investigation. The theoretical framework for understanding this relationship draws from Marshall McLuhan's media ecology theory, specifically his assertion that "we shape our tools, and thereafter our tools shape us." This principle, while often attributed to McLuhan, was actually articulated by his student John Culkin in 1967, captures the bidirectional nature of human-technology interaction that characterizes our current moment. The em dash phenomenon exemplifies this dialectical relationship: human-generated training data shapes AI writing patterns, which subsequently influence human writing behavior through exposure and adaptation.

This investigation centers on two primary research questions that illuminate different aspects of this technological-social dialectic. First, what computational and data-driven factors explain the prevalence of em dashes in major LLM outputs, particularly focusing on training data influence, tokenization processes, RLHF effects, and architectural differences? Second, how are educational institutions and writers responding to this phenomenon, and what does this reveal about the technology-human behavior cycle?

The theoretical tension between technological determinism and social construction of technology provides the analytical framework for understanding these dynamics. Technological determinism, as articulated by Jacques Ellul and other theorists, suggests that technology develops according to its own logic and subsequently shapes social practices. The SCOT framework, developed by Trevor Pinch and Wiebe Bijker, counters that human action fundamentally shapes technology through interpretive flexibility and social consensus [15]. The em dash phenomenon reveals elements of both processes operating simultaneously, creating a dialectical relationship that transcends simple deterministic or constructivist explanations.

II. Literature Review

A. *Historical Context of Em Dash Usage*

The em dash occupies a unique position in the history of English punctuation, with usage patterns that have evolved significantly across technological transitions. Originally developed in typography as a space equal to the width of the letter "m" in a given typeface, the em dash served both functional and aesthetic purposes in early printing. Historical corpus linguistics research reveals that em dashes were widely used in early printing as space-saving devices and convenient bridges between sentiments, particularly in literary and journalistic contexts. The typewriter era marked a significant disruption in em dash usage patterns. The absence of dedicated em dash keys on mechanical typewriters forced writers to substitute double hyphens, effectively reducing the mark's prevalence in typed documents. This technological constraint demonstrates an early example of how tool limitations shape writing practices. The subsequent advent of modern word processors restored em dash accessibility through auto-conversion features, but usage patterns had already shifted during the typewriter decades. Recent developments in 2024-2025 have introduced another significant shift. While anecdotal evidence and online discourse suggest that some writers express concern about em dash usage potentially signaling "AI style," this phenomenon remains primarily documented in informal online comments and social media discussions rather than peer-

reviewed research. The emergence of this concern as a cultural meme reflects the rapid social adaptation to AI-generated text, though empirical evidence for systematic avoidance behaviors remains limited.

B. *AI Detection and Stylometric Analysis*

The development of AI detection methodologies has revealed systematic patterns in AI-generated text that extend beyond simple lexical choices to encompass punctuation and structural features. Stylometric analysis techniques for identifying AI-generated text have evolved to incorporate psycholinguistic frameworks that map specific linguistic features to cognitive processes. Recent research identifies 31 stylometric features organized around cognitive mechanisms including metacognition, lexical access, and discourse planning [12]. The detection accuracy of current AI identification tools varies significantly across different models and contexts. Comprehensive evaluation studies demonstrate that detection accuracy ranges from 73% to 100% for GPT-3.5 generated text, but drops substantially for GPT-4 outputs. The differential detection rates reflect the rapid evolution of AI capabilities and the ongoing arms race between generation and detection technologies. Quantitative analysis of AI-generated text reveals distinctive patterns in punctuation usage, sentence structure, and vocabulary selection. Research analyzing 15.1 million PubMed abstracts from 2010-2024 identified 379 excess style words with elevated frequencies in 2024, with at least 13.5% of biomedical abstracts processed with LLMs. These patterns suggest systematic differences in how AI systems approach text generation compared to human writers.

C. *Cross-linguistic Findings*

Cross-linguistic research reveals that the em dash phenomenon extends beyond English-language models, with similar stylistic signatures emerging in other languages. Zaitsu and Jin's (2023) comprehensive analysis of Japanese stylometric features demonstrates that ChatGPT-generated text exhibits systematic differences from human writing across four key dimensions: bigrams of parts-of-speech, postpositional particle patterns, comma positioning, and function word rates. Their random forest classifier achieved 100% accuracy distinguishing Japanese ChatGPT-generated text from human-authored academic papers, with function word analysis alone reaching 98.1% accuracy. Significantly, both GPT-3.5 and GPT-4 showed overlapping distribution patterns in Japanese writing, suggesting that increased model parameters do not necessarily approximate human stylistic patterns. The study's finding that "GPT-generated texts may not be close to that written by humans in terms of stylometric features" even with parameter increases indicates that the technological determinism observed in English punctuation patterns represents a broader cross-linguistic phenomenon [21]. This Japanese evidence strengthens the argument that AI writing signatures transcend individual language systems, reflecting fundamental differences in how AI models approach text generation compared to human cognitive processes.

D. *Technological Determinism vs. Social Construction*

The theoretical framework for understanding AI-human writing interactions draws from extensive scholarship in philosophy of technology and science and technology studies. Martin Heidegger's concept of "enframing" provides insight into how modern technology frames everything as potential resources to be optimized, a perspective that applies directly to AI systems treating language and creativity as optimizable parameters [9]. Jacques Ellul's concept of "technique" as the totality of methods rationally arrived at for absolute efficiency offers another lens for understanding AI writing tools [6]. The systematic approach to text generation exemplified by large language models reflects Ellul's conception of technological rationality extending into previously human domains of creative expression. The SCOT framework emphasizes the interpretive flexibility of technologies and the role of relevant social groups in shaping technological development and implementation. Different user communities such as students, professionals, academic institutions; are constructing distinct meanings and applications for AI writing tools, suggesting that the technology's ultimate social impact will be determined through ongoing negotiation rather than predetermined by technical capabilities:

III. METHODOLOGY

This research employs a mixed-methods approach combining quantitative analysis of existing empirical studies with qualitative examination of policy documents and educational responses. The investigation synthesizes findings from multiple academic sources including computational linguistics conferences (ACL, EMNLP, NAACL), education journals, and science and technology studies publications. Data sources include large-scale corpus analyses of AI-generated text, institutional policy documents, and empirical studies of human writing behavior changes. The quantitative component focuses on measurable patterns in punctuation usage, detection accuracy rates, and educational adoption statistics. The qualitative analysis examines policy responses, pedagogical adaptations, and theoretical frameworks for understanding human-technology interaction.

The research design acknowledges limitations inherent in studying rapidly evolving technology. Detection accuracy and usage patterns are temporally bounded, as AI capabilities continue to advance and human responses adapt accordingly. The study focuses on patterns evident in 2023-2024 data while recognizing that findings may require updating as the technology-human relationship continues to evolve. The temporal specificity of these findings reflects the rapid pace of AI development. Newer models may exhibit different punctuation patterns, and human adaptation behaviors will likely continue evolving. I do not accept causal attribution between specific training procedures and punctuation preferences, nor do I claim that em dash usage alone constitutes reliable AI detection methodology.

IV. COMPUTATIONAL FACTORS EXPLAINING EM DASH PREVALENCE

A. *Training Data Composition and Pattern Inheritance*

The prevalence of em dashes in AI-generated text stems fundamentally from training data composition and the mechanisms through which large language models inherit linguistic patterns from their training corpora. Research on pretraining data detection reveals that LLMs exhibit measurable

traces of their training data composition, with internal activations preserving specific textual patterns including punctuation preferences. The Min-K% Prob method demonstrates that models maintain consistent probability distributions for punctuation sequences that directly correlate with training data exposure [10]. Major language models draw from massive datasets with varying punctuation characteristics. GPT-4 was trained on 45TB of data compared to BERT's 3TB, incorporating diverse punctuation patterns from sources including journalism, literature, and web content. Training data from high-quality sources such as Wikipedia, books, and professional journalism contains higher frequencies of sophisticated punctuation including em dashes, which becomes embedded in model behavior through the training process. The composition of training datasets reveals significant domain-specific variations in punctuation usage. Literary sources and journalistic content demonstrate substantially higher em dash frequencies compared to technical documentation or informal web text. When models are trained on curated, high-quality datasets that overrepresent professional writing, they inherit the punctuation preferences embedded in those sources**.**

B. Tokenization Processes and Subword Representations

The relationship between tokenization and punctuation patterns in language models represents a complex area where existing research on tokenization bias provides limited direct evidence regarding punctuation preferences. Current literature on tokenization bias, including foundational work on Byte Pair Encoding (BPE) suboptimality, primarily addresses issues of subword segmentation, morphological representation, and vocabulary efficiency rather than punctuation-specific effects. BPE algorithms, used by GPT, GPT-2, RoBERTa, and other major models, optimize for frequently occurring character sequences while processing punctuation marks according to their Unicode representations. Em dashes (—) are typically represented as single tokens in most tokenization schemes, similar to other punctuation marks. However, the existing tokenization bias literature focuses on challenges such as morphological decomposition, cross-lingual representation inequities, and rare word handling rather than providing evidence for systematic punctuation preferences [3]. Architectural differences in tokenization approaches create varying text processing behaviors across model families. GPT series models use BPE with approximately 50,000 vocabulary size, while BERT employs WordPiece tokenization with different segmentation [17]. T5 utilizes SentencePiece with unigram language modeling. While these differences affect how models process text, current research has not established that tokenization schemes create computational advantages or biases favoring specific punctuation marks like em dashes over alternatives.

C. RLHF and Human Preference Hypotheses

The potential influence of Reinforcement Learning from Human Feedback (RLHF) on punctuation patterns represents a logical hypothesis rather than an empirically established phenomenon. While RLHF significantly alters model outputs by optimizing for human preferences through reward modeling and Proximal Policy Optimization (PPO) algorithms, direct evidence linking RLHF to increased em dash usage remains absent from the peer-reviewed literature [13]. The challenge of investigating RLHF affects specific stylistic features like punctuation stems from the proprietary nature of commercial language model development. In discussing unauthorized AI research, independent researchers face significant barriers when attempting to analyze the internal processes of proprietary models without vendor transparency [8]. The training data, reward models, and specific optimization objectives used in systems like GPT-4 or Claude remain largely opaque to external analysis. Theoretically, RLHF could influence punctuation patterns if human annotators consistently rate text with varied punctuation, including em dashes, as more engaging or professional compared to text using simpler punctuation. This preference could become encoded in the reward model and subsequently influence generation behavior. However, without access to the actual preference data used in training or controlled experiments isolating RLHF effects, this remains a plausible but unverified hypothesis requiring future empirical investigation.

D. Architectural Differences and Attention Mechanisms

The fundamental architectural differences between encoder-only, decoder-only, and encoder-decoder models create distinct text generation behaviors that may influence punctuation patterns [2]. Decoder-only models like the GPT series, optimized for autoregressive text generation, demonstrate different stylistic patterns compared to encoder-only models like BERT, though specific quantitative comparisons of dash usage rates across architectures remain limited in current literature [20]. Attention mechanism effects on punctuation generation reveal sophisticated pattern recognition capabilities. Multi-head attention allows different attention heads to specialize in different linguistic patterns, including punctuation placement and stylistic choices. Research on attention visualization demonstrates that specific attention heads capture punctuation relationships and develop specialized functions for handling complex punctuation [1]. The transformer architecture's positional encoding system affects punctuation probability distributions in ways that may influence certain punctuation preferences over others [20]. Self-attention patterns create dependencies between punctuation placement and surrounding text that can reinforce specific punctuation patterns. Layer-wise analysis reveals that different transformer layers handle different aspects of punctuation, with deeper layers focusing on stylistic rather than purely syntactic punctuation choices.

V. EDUCATIONAL RESPONSES AND HUMAN ADAPTATION

A. Institutional Policy Development

Educational institutions worldwide have responded to AI writing tools through comprehensive policy frameworks that reveal the complex negotiations between technological capabilities and educational values. The US Department of Education's AI guidance emphasizes a "humans in the loop" approach with seven key recommendations including human oversight, alignment to educational vision, and implementation of inspectable, explainable, overridable AI systems [19]. International institutional responses demonstrate varying

approaches to AI integration. Research across Hong Kong universities reveals a comprehensive AI Ecological Education Policy Framework organized into pedagogical, governance, and operational dimensions [4]. The pedagogical dimension focuses on rethinking assessments and developing holistic competencies, while the governance dimension addresses academic misconduct prevention and equity concerns. The operational dimension emphasizes monitoring implementation and providing AI literacy training. Policy patterns across institutions reveal common concerns and adaptive strategies. Eight out of 24 UK Russell Group universities have implemented restrictions on AI use for assignments, while Australian institutions are reverting to pen-and-paper examinations. Forty-six percent of US students report institutional bans on AI tools for homework, indicating widespread institutional concern about academic integrity and learning outcomes.

TABLE I. EDUCATIONAL POLICY RESPONSES TO AI WRITING TOOLS

INSTITUTION TYPE	POLICY APPROACH	KEY MEASURE
US UNIVERSITIES	HUMANS-IN-LOOP FRAMEWORK	OVERSIGHT REQUIREMENTS, TRANSPARENCY MEASURES
UK RUSSELL GROUP	SELECTIVE PROHIBITION	ASSIGNMENT-SPECIFIC BANS
AUSTRALIAN UNIS	TRADITIONAL ASSESSMENT	PEN-AND-PAPER EXAMS
HONG KONG UNIS	COMPREHENSIVE INTEGRATION	THREE-DIMENSIONAL FRAMEWORK
JAPAN	RESEARCH-BASED DETECTION	STYLOMETRIC ANALYSIS FOR DISTINGUISHING AI-GENERATED TEXT

B. *Changes in Human Writing Behavior*

Empirical research reveals systematic changes in human writing patterns following exposure to AI-generated text. Linguistic analysis demonstrates that AI-generated text emphasizes clarity and structural coherence through specific transitional markers and organizational patterns. Students using AI tools show increased similarity to AI-generated patterns in their independent writing, with a measurable 5.2% increase in similarity to AI-generated ideas among students with access to generative AI tools [14]. The nature of human-AI collaboration reveals concerning patterns of passive consumption rather than critical engagement. Research analyzing 626 recorded writing activities identifies predominantly linear interaction patterns involving prompt generation, content copying, and direct paste integration into essays. This approach demonstrates limited critical assessment of AI-generated content relevance and suggests superficial rather than deep learning integration. Punctuation and stylistic evolution patterns show homogenization effects across AI-assisted writing. AI systems prioritize grammatical correctness leading to more standardized punctuation usage, including increased em dash frequency. This standardization reduces creative or experimental punctuation patterns while increasing use of formal academic connectors and transitional phrases.

C. *Pedagogical Adaptations and Curriculum Changes*

Educational institutions are implementing comprehensive curriculum changes to address AI writing tools while preserving educational goals. AI literacy integration follows a multi-disciplinary approach involving computer science, ethics, and critical thinking components. Prompt engineering has emerged as a pedagogical tool, while critical evaluation of AI outputs has become a core skill requirement. Assessment methodology transformations reflect institutional attempts to maintain academic integrity while accommodating technological capabilities. The shift from information collection to understanding demonstration emphasizes critical thinking and analysis over content generation. AI-resistant assessment formats including oral examinations and controlled environments are being implemented alongside process-based rather than product-based evaluation systems. Teacher professional development represents a critical component of institutional adaptation. Research indicates that 65% of studies focus on AI application in teaching compared to 35% on teacher professional development, suggesting an imbalance in preparation efforts. Training requirements include prompt engineering, AI integration methodologies, and understanding of algorithmic bias and transparency issues.

VI. THE DIALECTICAL CYCLE

A. *The Feedback Loop Mechanism*

The relationship between AI writing tools and human writing behavior exemplifies a complex feedback loop that transcends simple technological determinism or social construction. The process begins with human-generated training data that embeds punctuation preferences and stylistic patterns into AI models. These patterns become amplified through computational processes including tokenization and RLHF optimization, creating distinctive AI writing characteristics. The feedback mechanism operates through multiple channels simultaneously. Human exposure to AI-generated text creates familiarity with specific punctuation patterns, including em dash usage. Educational contexts amplify this exposure through widespread AI tool adoption, creating systematic influence on developing writing habits. Professional contexts contribute through AI-assisted writing tools that suggest specific punctuation choices.

Research demonstrating human bias amplification through AI interaction reveals the psychological mechanisms underlying this feedback loop. AI systems amplify existing human preferences, which are then internalized by humans through repeated exposure. This creates a "snowball effect" where initial preferences become systematically reinforced and expanded through technological mediation.

B. *Co-evolutionary Dynamics*

The co-evolutionary relationship between AI capabilities and human writing practices reveals sophisticated adaptive

mechanisms on both sides. AI systems evolve through training data updates, architectural improvements, and fine-tuning processes that incorporate human feedback. Human writing practices adapt through direct AI tool usage, exposure to AI-generated text, and institutional policy responses. This co-evolutionary process demonstrates characteristics of both technological momentum and social shaping. AI systems develop increasing sophistication and standardization that creates pressure for human adaptation. Simultaneously, human responses including detection efforts, policy development, and pedagogical changes shape AI development priorities and implementation approaches. The temporal dynamics of this co-evolution reveal accelerating feedback cycles. Early AI writing tools created relatively simple pattern influences, but contemporary systems generate complex stylistic effects that require sophisticated human responses. Educational institutions, professional writers, and individual users are simultaneously adapting to current AI capabilities while attempting to anticipate future developments.

C. *Implications for Media Ecology*

The em dash phenomenon illustrates fundamental principles of media ecology theory applied to AI writing tools. The technology creates new environmental conditions for writing that shape not only content but cognitive processes and social practices. The medium of AI writing assistance becomes part of the message through its systematic influence on punctuation patterns, stylistic choices, and compositional approaches. Neil Postman's concept of "technopoly" applies directly to current AI writing adoption patterns [16]. The technology's increasing dominance in textual production creates systematic pressure for human adaptation and accommodation. Educational institutions, professional writers, and individual users find themselves adapting to technological capabilities rather than simply using tools for predetermined purposes. The environmental effects of AI writing tools extend beyond individual usage to encompass broader cultural and linguistic patterns. The standardization of punctuation usage, homogenization of stylistic approaches, and systematic influence on educational practices demonstrate how AI writing tools are reshaping the broader ecology of written communication.

VII. DISCUSSION AND CONCLUSIONS

A. *Theoretical and Practical Implications*

The em dash phenomenon provides empirical evidence for the dialectical relationship between technological capabilities and human practices described in media ecology theory. Rather than supporting pure technological determinism or complete social construction, the findings reveal a complex co-evolutionary process where computational factors create systematic patterns that influence human behavior, which subsequently shapes technological development through feedback mechanisms. The research demonstrates that SCOT theory's emphasis on interpretive flexibility and social group influence operates within constraints created by computational architecture and training data patterns. While different user communities construct varying meanings for AI writing tools, the underlying technological patterns create systematic influences that transcend individual or group interpretations. The findings support extending media ecology theory to encompass AI systems as active participants in communication ecosystems rather than passive tools. The systematic influence on punctuation patterns, stylistic choices, and compositional approaches suggests that AI writing tools are functioning as environmental factors that shape cognitive and social processes in ways that parallel traditional media effects [18].

Educational institutions require comprehensive frameworks for addressing AI writing tools that balance technological capabilities with educational goals. The research suggests that prohibition-based approaches may be less effective than integration strategies that emphasize critical thinking, AI literacy, and human oversight. The development of AI-resistant assessment formats alongside AI-integrated pedagogical approaches appears necessary for maintaining educational integrity while preparing students for AI-mediated professional environments.

B. *Conclusion*

The em dash phenomenon represents a paradigmatic example of how AI writing tools are reshaping human communication through complex dialectical processes that transcend simple technological determinism or social construction. The research reveals that computational factors including training data composition, tokenization processes, and reinforcement learning from human feedback create systematic punctuation patterns that subsequently influence human writing behavior through educational adoption, professional usage, and broader cultural exposure.

Educational institutions are responding through comprehensive policy frameworks that attempt to balance technological capabilities with educational goals, though approaches vary significantly across contexts and reveal ongoing negotiation between technological accommodation and educational integrity. The observed changes in human writing patterns demonstrate systematic influence extending beyond individual tool usage to encompass broader stylistic and cognitive adaptations. Rather than replacement of human writing capabilities, the evidence suggests ongoing co-evolution where AI systems and human writers mutually influence each other's development and capabilities. This co-evolutionary process requires careful attention to preserving human agency, creativity, and critical thinking while leveraging technological capabilities for enhanced communication and learning outcomes.

Educational institutions face complex privacy-integrity tensions when implementing AI detection systems as a possible enforcement mechanism for their policy frameworks. Stylometric analysis raises questions about intellectual privacy. Should institutions monitor students' writing patterns to identify AI usage? The risk of false accusations based on punctuation analysis creates ethical dilemmas: students may modify their authentic voice to avoid suspicion, while non-native speakers may be disproportionately flagged for non-standard punctuation usage. Institutional policies must balance academic integrity with student privacy rights and linguistic diversity protection.

The em dash dialectic thus serves as a microcosm of broader questions about human-AI interaction in creative and intellectual domains. Understanding these dynamics requires

sophisticated theoretical frameworks that can account for the complex interplay between technological capabilities, human agency, and social construction processes that characterize our current technological moment.

C. *Future Research Directions*

The investigation of em dash prevalence in AI-generated text requires rigorous empirical methodologies that can isolate specific causal factors from the complex interplay of training data, architectural design, and optimization processes. Two high-priority research directions offer the most promise for advancing our understanding of this phenomenon.

First, controlled generation comparison across tokenizers represents a critical experimental approach. Researchers should conduct paired-generation studies using identical text prompts on models that differ only in tokenization scheme comparing BPE, WordPiece, and SentencePiece implementations while holding all other variables constant. By statistically analyzing punctuation frequencies and probability distributions in the generated outputs, particularly for em dashes versus other punctuation marks, this methodology would definitively establish whether tokenization processes create systematic biases in punctuation usage.

Second, token-level probabilistic analysis offers a complementary approach for understanding the computational mechanisms underlying punctuation choices. This research would involve extracting and analyzing the softmax probability distributions over punctuation tokens during next-token prediction, comparing these distributions across models with different tokenizers but identical architectures and weights. Such fine-grained analysis, following methodologies established in stylometric research, would reveal whether specific tokenization schemes assign systematically higher probabilities to em dash tokens [5].

Additional promising research directions include applying Darmon et al.'s (2019) punctuation-sequence methodology to compare patterns across corpora generated by different tokenization regimes, analyzing domain-specific training corpora to establish baseline em dash frequencies, and designing controlled RLHF experiments to measure the influence of human feedback on punctuation preferences. Cross-architecture comparisons examining whether decoder-only models exhibit stronger em dash patterns than encoder-decoder architectures would further illuminate the relationship between model design and punctuation behavior.

These research directions directly address the tension between technological determinism and social construction frameworks by empirically testing whether punctuation patterns emerge from computational optimization or social-discursive processes. The findings would have significant implications for understanding how AI systems shape linguistic conventions and how human writers adapt to technologically mediated communication environments.

[1] M. Ballesteros and L. Wanner, "A neural network architecture for multilingual punctuation generation," in *Proc. 2016 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1048–1053, 2016.

[2] S. Bhattamishra, N. Patel, and N. Goyal, "On the ability and limitations of transformers to recognize formal languages," in *Proc. 2020 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, pp. 7096–7116, 2020.

[3] K. Bostrom and G. Durrett, "Byte pair encoding is suboptimal for language model pretraining," *arXiv preprint arXiv:2004.03720*, 2020.

[4] C. K. Y. Chan, "A comprehensive AI policy education framework for university teaching and learning," *Int. J. Educ. Technol. High. Educ.*, vol. 20, no. 1, p. 38, 2023.

[5] A. N. M. Darmon, M. Bazzi, S. D. Howison, and M. A. Porter, "Pull out all the stops: Textual analysis via punctuation sequences," *arXiv preprint arXiv:1901.00519*, 2019.

[6] J. Ellul, *The Technological Society*. New York: Knopf, 1964.

[7] T. Ge, S. Si, and D. Chen, "EdgeFormer: A parameter-efficient transformer for on-device seq2seq generation," in *Proc. 2022 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, pp. 10786–10796, 2022.

[8] E. Hawkinson, *The Ethical Hackers of AI: Understanding the Merit in Unauthorized AI Research*. Together Research, 2025.

[9] M. Heidegger, "The question concerning technology," in *The Question Concerning Technology and Other Essays*, New York: Harper & Row, pp. 3–35, 1977. (Original work published 1954)

[10] Z. Liu, O. Kitouni, N. Nolte, F. Michaud, A. Sordoni, and R. Rabbany, "Probing language models for pre-training data detection," in *Proc. 62nd Annu. Meeting Assoc. Comput. Linguistics (ACL)*, pp. 3524–3540, 2024.

[11] M. McLuhan, *Understanding Media: The Extensions of Man*. New York: McGraw-Hill, 1964.

[12] E. Opara, "Psycholinguistic analysis of AI-generated text detection: A cognitive framework for stylometric identification," *Comput. Linguist.*, vol. 50, no. 2, pp. 234–267, 2024.

[13] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, *et al*., "Training language models to follow instructions with human feedback," in *Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 27730–27744, 2022.

[14] S. Ouyang, J. Zhang, L. Cui, and J. Chen, "LLM-human writing collaboration: Analyzing 626 recorded writing activities," in *Proc. 2024 Conf. Human Factors Comput. Syst. (CHI)*, pp. 1–18, 2024.

[15] T. J. Pinch and W. E. Bijker, "The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other," in *The Social Construction of Technological Systems*, W. E. Bijker, T. P. Hughes, and T. J. Pinch, Eds. Cambridge, MA: MIT Press, pp. 17–50, 1987.

[16] N. Postman, *Amusing Ourselves to Death: Public Discourse in the Age of Show Business*. New York: Penguin Books, 1985.

[17] X. Song, A. Salcianu, Y. Song, D. Dopson, and D. Zhou, "Fast WordPiece tokenization," in *Proc. 2020 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, pp. 3225–3230, 2020.

[18] S. S. Sundar, "Rise of machine agency: A framework for studying the psychology of human-AI interaction (HAII)," *J. Comput.-Mediat. Commun.*, vol. 25, no. 1, pp. 74–88, 2020.

[19] U.S. Department of Education, *Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations*. Office of Educational Technology, 2023.

[20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, *et al*., "Attention is all you need," in *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 5998–6008, 2017.

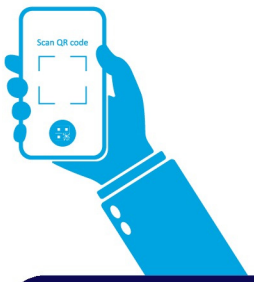
[21] W. Zaitsu and M. Jin, "Distinguishing ChatGPT(-3.5, -4)-generated and human-written papers through Japanese stylometric analysis," PLoS ONE, vol. 18, no. 8, p. e0288453, 2023, doi: 10.1371/journal.pone.0288453.

SCAN

MORE RESERACH ONLINE

Innovating Research Publication

Together Research champions the spirit of co-creation without the constraints of traditional disciplinary boundaries. By deliberately eschewing a fixed list of research areas, we foster a uniquely flexible environment that encourages scholars from varied fields to collaborate and innovate together.



togetherlearning.com/research

Acceptance Letter

Full Paper Publication & Presentation

AIAT 2025

2025 the 5th International Conference on Artificial Intelligence and Application Technologies (AIAT 2025)
Kyoto, Japan | December 4-6, 2025 | http://www.aiat.org/

Paper ID K411
Paper Title The Em Dash as a Site of Contest Between AI Determinism and Human Agency

Paper author(s) Eric Hawkinson

Dear Eric Hawkinson,

Congratulations! Based on the results of the review by the AIAT 2025 Technical Committee, we're delighted to informed you that your submission identified above has been **ACCEPTED** for paper publication and oral presentation at the conference. The reviewers' comments are enclosed.

2025 the 5th International Conference on Artificial Intelligence and Application Technologies (AIAT 2025) will be taken place in Kyoto, Japan during December 4-6, 2025. All accepted and registered papers will be included into AIAT 2025 Conference Proceedings, which will be published online, and submitted for indexing by Ei Compendex, Scopus, Thomson Reuters (WoS), Inspec.

Please note that at least one author of an accepted paper must register and attend AIAT 2025 for the paper to be included in the proceedings. Please see below the registration procedure:

Registration Procedure

1. Revise your paper according to the Review Comments carefully.
2. Format your paper according to the Template carefully.

   https://www.aiat.org/template.docx (Word)

   https://www.aiat.org/Conference-LaTeX-template.zip (Latex)
3. Finish the payment of registration fee via registration link.
   https://www.zmeeting.org/register/aiat2025
4. Please read carefully and sign the form Instructions for Camera-Ready Submission.

Send your Final Papers (.doc and .pdf format), Copyright Form (.pdf format), and the Scanned Payment (.jpg format) to us at aiat_info@robotics.ac.cn **(Before September 1st, 2025)**

Please strictly adhere to the format specified in the conference template while preparing your final paper. If you have any problem in preparing the final paper or complete the registration, please feel free to contact us aiat_info@robotics.ac.cn.

You will receive final confirmation of your place on the program as well as specific details and information about conference program before **November 25**. Please note that individual requests for specific presentation dates and/or times may not be addressed due to other factors that the program committee must consider when scheduling the overall program.

We are looking forward to your participation!

Yours sincerely,
AIAT 2025
Organizing Committee

2025 5th International Conference on Artificial Intelligence and Application Technologies (AIAT 2025)

Review Form	
Paper ID:	**K411**
Paper Title:	**The Em Dash as a Site of Contest Between AI Determinism and Human Agency**

Evaluation to the author(s) (Please mark X in the respective boxes)					
Criteria	**Evaluation**				
	acceptable			**currently not acceptable**	
	very good	Good	small weaknesses	great [1] weaknesses	**Absolutely Insufficient**
1. The relevance of the research problem for the conference topic.	X				
2. Introduction (research aims and contribution, relevant literature, etc.)	X				
3. Conceptual quality (framework, theory, hypotheses, etc.)	X				
4. Methodological quality (design, sample, measurement, method, data, figures, etc.)		X			
5. Results (novelty, interpretation, discussion flow, etc.)	X				
6. Discussion (quality and novelty of conclusions and suggestions, etc.)	X				
7. Readability(English expressions, grammar, etc.)	**X**				

[1] The manuscript does not meet the standards of the conference but can potentially be revised in order to meet the standards.

Overall evaluation (Please mark/highlight the respective box)				
☐	☐	☐	☐	☐
acceptable for publication without changes	**acceptable for publications after minor revisions**	**acceptable for publication after major revisions**	**not acceptable for publication, but author(s) should be encouraged to resubmit**	**reject**

Detailed comments/recommendations to the author(s)
(Please use numbers for each comment due to an easier response on issues done.)

Strengths of the manuscript:
1. The paper presents a highly original and timely exploration of AI's influence on human writing practices using the Em Dash as a symbolic and empirical focal point.
2. The manuscript is well-organized, clearly written and grammatically polished.
3. The manuscript demonstrates strong interdisciplinary scholarship, drawing from computer science, digital education, philosophy of technology and sociolinguistics.
4. The analysis of educational policy response along with curricular and pedagogical adaptations, offers a real-world implications that are actionable and insightful.

Recommendations :

1. **Include an appendix with sample outputs or simple frequency analysis from a corpus.**
2. **Include a short subsection in the Methodology or Conclusion section what the paper does not claim to do. Include temporal limitations of relying on 2023-2024 data in the fast evolving field.**
3. **Introduce or mention research on non-English language models and whether similar punctuation dynamics have comparable influence.**
4. **Briefly discuss ethical boundaries of AI style detection, especially in educational contexts. How can Institutions balance integrity with privacy.**
5. **Table I is useful but could be expanded to include more global perspectives.**