No. 25 ISSN 2450-0402

Discourses on Culture

Editor-in-chief: **Iga Maria Lehman**

Wydawnictwo Społecznej Akademii Nauk

Łódź 2026

No. 25 ISSN 2450-0402

Discourses on Culture

Editor-in-chief: **Iga Maria Lehman**

Łódź 2026

Wydawnictwo
Społecznej Akademii Nauk

The electronic version is the primary version of the journal

http://dyskursy.san.edu.pl

All the articles published in the journal are subject to reviews.

Editor-in-chief: **Iga Maria Lehman**

Adjustment: Julia Żak

Text design and typesetting: Studio Grafpa, www.grafpa.pl

Cover design: Studio Grafpa, www.grafpa.pl

© copyright Społeczna Akademia Nauk and Authors

under the Creative Commons License CC BY-NC-ND

https://creativecommons.org/licenses/by-nc-nd/4.0/

ISSN 2450-0402

University of Social Sciences Publishing House

Sienkiewicza 9, 90-113 Łódź

e-mail: discourses@san.edu.pl

WYDAWNICTWO
SPOŁECZNEJ AKADEMII NAUK

Table of Contents

7 **Janne Tienari** Preface: Algorithm and Author: Retaining the Human Touch?

Articles

17 **Michał Szostak** Algorithmic Kitsch: AI-Generated Writing as Aesthetic Simulation in Managerial Practice

37 **Rania Za'rour, Abdel Rahman Mitib Altakhaineh** False Positive Risk in AI Detection of L2 Academic Writing: A Case Study

79 **Irena Vassileva, Mariya Chankova** GenAI and Authorship in Academia: The (Im)Possible Creator. An Integrative Review

111 **Adrian Holliday** Working with ChatGPT: Engaging with the Real World and Our Own Academic Practice

135 **Jagat Bahadur Kunwar** Authenticity and Authorship Beyond the Algorithm: Action, Institution, and Epistemic Responsibility under Generative AI

165 **Ken Hyland** Rethinking AI in Writing Pedagogy: From Fluency to Rhetorical Competence

190 **Rahmad Nuthihar, Dadang S. Anshori, Isah Cahyani, Andoyo Sastromiharjo** Making Students "Safe" Writers with the SAFE Model in the Age of AI

221 **Djalel Baghzou, Assala Belsem Bouameur, Michał Szostak, Mourad Mahmoudi** From Mental Concept to Algorithmic Exemplification: Rethinking the Aesthetic Situation in the Age of Generative AI

245 **Muhammad Erwan Saing** Algorithmic Storytelling and Vernacular Folklore: Rewriting the Nenek Pakande Monster in the Age of Generative AI

279 **Binh Quoc Duong, Ha Thu Le, Minh Thi Tuyet Nguyen** Visual Politics in the Age of Algorithms: A Four-Dimensional Model of Meaning, Visibility, Affect, and Legitimacy

DOI: 10.2478/doc-2026-0001

This is an open access article licensed under the Creative Commons AttributionNonCommercial-NoDerivatives 4.0 International (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Preface

Algorithm and Author: Retaining the Human Touch?

Article history:

Received **10 June 2026**

Available online **23 June 2026**

I am honored to write a preface to the special issue on *Between Algorithm and Author: Rethinking Writing, Argumentation, and Voice in the Age of AI.* This unique collection of contributions could not be timelier and more relevant. Interrogating the shifting boundaries between algorithm and author, machine and human, and formula and originality is overdue as we experience generative artificial intelligence (Gen AI) disrupting our academic work and lives. Taking a step back to think through, and to critically discuss, crucial issues such as writing and authorship is a worthy endeavor when everything seems to be changing so rapidly. Turning the focus to how AI challenges our conceptualizations of authenticity and originality is very welcome. It is crucially important to debate

what we as human writers and authors can do not only to cope but to prosper in the age of AI. This cuts to the core of our craft as academic scholars.

What is happening to writing and authorship reflects the world that is increasingly complex. We experience it as exceedingly fast, ambiguous, uncertain, unpredictable, and full of tensions. It keeps surprising us. A lot of this is down to artificial intelligence as increasingly sophisticated AI systems, models, and tools are at our disposal. Gen AI is taking the world by storm, and it challenges our assumptions and habits, among other things, about what writing means and what authorship is. This is steered by immensely powerful forces. Merchants of AI products and services enchant us with their magical performances, construing an inevitable trajectory of progress (Tienari & Khoreva, 2025). They portray critical views on AI as outmoded (Vesa & Tienari, 2022).

Academia is experiencing an "exogeneous shock" with Gen AI (Krammer, 2025) and "no matter which angle is considered, and whether we like it or not, AI systems are here to stay" (Kulkarni et al., 2024, p. 207). In these conditions, outsourcing academic writing to machines seems to be a serious option for many. This is neither coincidental nor surprising. The need to adhere to strict writing conventions has become so ritualistic and widespread that in principle "one need not be a scholar to publish. Instead, one needs to learn and perfect the routines that lead to publication" (Rasheed & Priem, 2020, p. 160). Concerns about lack of authenticity, reflexivity, and dialogue in scholarly writing predate the rise of Gen AI. Saying something original and meaningful has become secondary to formulaic and instrumentalist writing that targets getting published in the right journals (Alvesson & Gabriel, 2013; Harley, 2019; Kiriakos & Tienari, 2019; Lehman & Tienari, 2025).

Many resist uncritical adoption of AI in academia, working to retain our craft and community (Bechky & Davis, 2025) and learning (Tienari, 2025), and to safeguard our shared critical thinking skills (Larson et al., 2024). Scholars also raise concerns about what Gen AI is doing to academic knowledge production (Cornelissen, 2026; Lindebaum & Fleming, 2024) and highlight the threat that it poses to our epistemic agency as human creators of that knowledge (Lindebaum et al., 2025). Critical voices call for our reflexivity and responsibility as humans

to engage in creative, contextual, committed, and ethical research, writing, and authorship.

This seems increasingly difficult as more magic awaits around the corner (Tienari & Khoreva, 2025). Agentic AI is all the rage as AI is let loose to do its thing without human interception and steering. AI agents perceive their environment, set goals, plan, reason, and act independently of humans, and accomplish complex tasks with minimal intervention. We can expect something even more perplexing and possibly disastrous for writing and authorship in the not-so-distant future. In the worst-case scenario (or best, depending on the viewpoint), no humans are needed for producing academic knowledge. AI can do the research, write the paper, submit it, review it, make an editorial decision, publish the paper, read it, cite it… well, AI can in principle do everything that we humans can, only much quicker and more "efficiently." Whether this makes any sense at all is another matter. Yet the low-quality AI-augmented submissions that are already overwhelming journal editors (Gartenberg et al., 2026) can be multiplied many times over until the global academic publishing system reaches a breaking point.

However, there is still some time for us to react and challenge uncritical adoption of AI. For those of us who love writing (Kiriakos & Tienari, 2018) this involves arguing for human academic writing and authorship as something to be cherished and cared for. The contributions in this special issue showcase the multiplicity of ways we can approach and study writing and authorship creatively under conditions of AI hype.

Michał Szostak explores AI-generated managerial writing in organizations and discusses the notion of "algorithmic kitsch" to capture how it lacks epistemic depth and authentic subjectivity while simulating meaning through aesthetic coherence and persuasive form. Turning to academia, Rania Za'rour and Abdel Rahman Mitib Altakhaineh show how second language academic writing, due to its conventionalized nature, may face heightened risk of false positive AI flagging in AI-text detectors.

Irena Vassileva and Mariya Chankova extend the discussion to authorship under conditions of Gen AI with its convincing imitations of human natural language and considers how the changing position of the author relates to

acknowledging the intellectual contributions of other scholars. Looking at authorship from a different angle, Adrian Holliday elucidates how when prompted, ChatGPT reflects discursive and ideological struggles in the field of intercultural communication studies by privileging unreflective positivism over critical understandings of different social settings and events. Jagat Bahadur Kunwar, in turn, explores whether Gen AI can be considered an authentic author and argues that while it can perform linguistic labor, support linguistic work, and approximate institutional markers of authorship, Gen AI cannot initiate linguistic action or assume epistemic responsibility and thereby it is unable to satisfy the conditions of authentic authorship.

Focusing on university students, Ken Hyland explores the ways in which Gen AI tools reshape students' academic writing and argues for viewing such tools as a supportive tutor rather than a proxy author in new writing contexts where human creativity and algorithmic fluency intersect. Rahmad Nuthihar et al. study university students' opinion-writing skills and highlight their difficulties in dealing with writing conventions and argumentative reasoning in the age of AI.

Taking us to the world of art, Djalel Baghzou et al. look at artistic creation and argue that Gen AI functions as a technical mediator, translating an artist's internal "mental concept" into a publicly accessible artefact whilst reconfiguring the ontic status of the artwork and the agency of the creator. With the example of a folklore figure, Muhammad Erwan Saing explores the crucial question of how algorithmic text generation and AI-mediated rewriting affect vernacular narrative tradition by reshaping narrative voice and cultural memory.

Finally, extending the discussion to the realm of politics, Binh Duong et al. address visual politics in the age of algorithms and frame it as a discursive and infrastructural process shaped by hybrid human-machine assemblages, showing how visibility operates as a provisional source of authority, accelerating affective and legitimacy effects beyond human deliberation.

Together, contributions in the special issue elucidate the multitude of issues at hand when we consider writing, authorship, argumentation, and voice under conditions of Gen AI. They remind us how we cannot sit idly as boundaries between algorithm and author, machine and human, formula and originality continue to shift and perhaps even change for good, favoring algorithm,

machine, and formula, if we stay quiet. We must take a stand, as individuals and collectives, and protect what we hold dear. In doing this, I think, it is important to note that to nurture human writing and authorship in the face of GenAI, we must also protect and save human reading. Writing and reading are two sides of the same coin as we write to be read, and we read so that we can write more compellingly.

When Gen AI threatens to take over everything in academic work, it is crucial to discuss what we as human writers and readers can do to protect and develop our craft (Bell & Willmott, 2020). This calls for action and contributions to this special issue offer important ideas and insights for envisioning what the issues at hand are and can be. As craftspeople of knowledge production—as researchers and educators, and as authors, reviewers, and editors—we have responsibilities that AI cannot handle. Our exposure to AI may be inevitable, but we can influence how exactly human-AI relations and interaction unfold. So, let's not lose the human touch in writing and authorship. It is still very much needed.

Janne Tienari
Management and Organisation, Hanken School of Economics, Finland
jtienari@hanken.fi
ORCID ID: 0000-0002-2605-2760

Biography

Janne Tienari is Professor of Management and Organisation at Hanken School of Economics, Helsinki, Finland. His research and teaching interests include strategy work; people and AI; gender and diversity; identities, branding, media, and social media; and changing academia from a critical perspective. Tienari's work is published in leading international journals. He has served as Senior Editor of Organisation Studies and as Associate Editor of Organisation.

References

Alvesson, M., & Gabriel, Y. (2013). Beyond formulaic research: In praise of greater diversity in organizational research and publications. *Academy of Management Learning & Education*, *12*(2), 245–263. DOI: 10.5465/amle.2012.0327.

Bechky, B. A., & Davis, G. F. (2025). Resisting the algorithmic management of science: Craft and community after generative AI. *Administrative Science Quarterly*, *70*(1), 1–22. DOI: 10.1177/00018392241304403.

Bell, E., & Willmott, H. (2020). Ethics, politics and embodied imagination in crafting scientific knowledge. *Human Relations*, *73*(10), 1366–1387. DOI: 10.1177/0018726719876687.

Cornelissen, J. (2026). The AI con: The commodification of theory and the new and improved credibility crisis. *Organization Studies*, in press. DOI: 10.1177/01708406261450467.

Gartenberg, C., Hasan, S., Murray, A., & Pierce. L. (2026). More versus better: Artificial intelligence, incentives, and the emerging crisis in peer review. *Organization Science*, in press. DOI: 10.1287/orsc.2026.ed.v37.n3.

Harley, B. (2019). Confronting the crisis of confidence in management studies: Why senior scholars need to stop setting a bad example. *Academy of Management Learning & Education*, *18*(2), 286–297. DOI: 10.5465/amle.2018.0107.

Kiriakos, C. M., & Tienari, J. (2018). Academic writing as love. *Management Learning*, 49(1), 263–277. DOI: 10.1177/1350507617753560.

Krammer, S. M. (2025). Is there a glitch in the matrix? Artificial intelligence and management education. *Management Learning*, *56*(93), 367–388. DOI: 10.1177/13505076231217667.

Kulkarni, M., Mantere, S., Vaara, E., van den Broek, E., Pachidi, S., Glaser, V. L., Gehman, J., Petriglieri, G., Lindebaum, D., Cameron, D. L., Rahman, H. A., Islam, G., & Greenwood, M. (2024). The future of research in an artificial intelligence-driven world. *Journal of Management Inquiry*, *33*(3), 207–229. DOI: 10.1177/10564926231219622.

Larson, B. Z., Moser, C., Caza, A., Muehlfeld, K., & Colombo, L. A. (2024). Critical thinking in the age of generative AI. *Academy of Management Learning & Education*, *23*(3), 373–378. DOI: 10.5465/amle.2024.0338.

Lehman, I. M., & Tienari, J. (2025). What do you mean? Linguistic sensitivity and relational reflexivity in scholarly writing. *Organization*, *32*(3), 464–475. DOI: 10.1177/13505084241233956.

Lindebaum, D., & Fleming, P. (2024). ChatGPT undermines human reflexivity, scientific responsibility and responsible management research. *British Journal of Management*, *35*(2), 566–575. DOI: 10.1111/1467-8551.12781.

Lindebaum, D., Nolan, E, Ashraf, M., Islam, G., & Ramirez, M. (2025). The transformation of epistemic agency and governance in higher education through Large Language Models–toward a future of organized immaturity. *Organization Studies*, in press. DOI: 10.1177/01708406251392002.

Rasheed, A. A., & Priem, R. L. (2020). Exchanges: "An A is an A": We have met the enemy, and he is us! *Academy of Management Perspectives*, *34*(1), 155–163. DOI: 10.5465/amp.2019.0116.

Tienari, J. (2025). Learning about artificial intelligence? Cluelessness, courage, and magic in facing a moving target. *Management Learning*, in press. DOI: 10.1177/13505076251361431.

Tienari J., & Khoreva, V. (2025). Through the magic lens: Management, language, and AI. *Discourses on Culture*, *24*(1), 45–71. DOI: 10.2478/doc--2025-0013.

Vesa, M., & Tienari, J. (2022). Artificial intelligence and rationalized unaccountability: Ideology of the elites? *Organization*, *29*(6), 1133–1145. DOI: 10.1177/1350508420963872.

Articles

DOI: 10.2478/doc-2026-0002

This is an open access article licensed under the Creative Commons AttributionNonCommercial-NoDerivatives 4.0 International (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Michał Szostak

Institute for Management Research, Civitas University, Warsaw, Poland

michal.szostak@civitas.edu.pl

ORCID ID: 0000-0002-7774-2964

Algorithmic Kitsch: AI-Generated Writing as Aesthetic Simulation in Managerial Practice

Article history:

Received	**21 March 2026**
Revised	**20 April 2026**
Accepted	**22 April 2026**
Available online	**23 June 2026**

Abstract: The proliferation of generative artificial intelligence (AI) in managerial practice has introduced a new paradigm of organisational writing that is fluent,

coherent, and rhetorically effective, yet increasingly detached from human intentionality and lived experience. This article conceptualises this phenomenon as algorithmic kitsch, defined as AI-generated discourse that simulates meaning through aesthetic coherence and persuasive form while lacking epistemic depth and authentic subjectivity. Drawing on theories of kitsch from aesthetic philosophy and their extension into management studies, the paper situates AI-generated texts within a broader trajectory of organisational communication characterised by standardisation, emotional templating, and symbolic overproduction. It argues that generative AI does not merely assist managerial writing but intensifies pre--existing tendencies toward aesthetic simplification and rhetorical cliché, thereby transforming argumentation into a form of patterned plausibility. The article outlines the mechanisms of algorithmic kitsch (i.e., aesthetic smoothing, semantic flattening, and affective pre-structuring) and discusses their implications for epistemic integrity, ethical responsibility, and the erosion of authentic managerial voice. From a humanistic management perspective, the paper calls for greater consciousness in the use of AI-generated texts and renewed attention to aesthetic literacy and critical reflexivity in AI-mediated managerial communication practices.

Keywords: management aesthetics, algorithmic kitsch, AI-mediated organisational communication, human-AI authorship, organisational discourse standardisation

Introduction

The diffusion of generative AI has significantly transformed writing practices across professional, academic, and organisational contexts. These AI-powered tools can produce texts that are grammatically correct, stylistically consistent, and rhetorically persuasive, often indistinguishable from those authored by humans at first glance. This technological shift raises fundamental questions about the nature of writing, authorship, and meaning-making in contemporary organisations. As scholars increasingly note, AI-generated discourse challenges not only technical aspects of communication but also its epistemological and ethical foundations (Bender et al., 2021; Floridi & Chiriatti, 2020).

Within management and organisational studies, communicating (including in writing) is not merely a neutral process for transmitting information and gaining feedback; it constitutes a central mechanism of organising, sensemaking, and leadership. Managerial texts (strategies, reports, mission statements, or ESG disclosures) perform symbolic and aesthetic functions, shaping how organisations are perceived and how meaning is constructed internally and externally (Barry & Elmes, 1997; Boje, 2001). From the perspective of management aesthetics, such texts can be understood as aesthetic artefacts that embody particular forms, styles, and sensibilities (Strati, 1999). Consequently, the increasing reliance on AI in producing these texts invites a critical examination not only of their functionality but also of their aesthetic and cultural qualities.

One particularly useful lens for analysing this transformation is the concept of kitsch. Traditionally associated with mass-produced art characterised by imitation, emotional simplification, and lack of originality, kitsch has been extensively discussed in aesthetic theory as a form of cultural expression that privileges immediacy and accessibility over depth and authenticity (Greenberg, 1985; Kulka, 1996; Scruton, 2011). More recent developments have extended the notion of kitsch into organisational contexts, where it captures forms of communication that are formulaic, emotionally appealing, yet substantively shallow and often evident in corporate narratives, branding, or managerial rhetoric (Kostera, 1997; Linstead, 2002; Linstead & Höpfl, 2000; Szostak, 2021, 2024a, 2024b, 2025). In this sense, kitsch is not merely a stylistic category but a mode of meaning production that reflects broader cultural and institutional dynamics.

Generative AI appears to amplify precisely those characteristics that define kitsch. By design, large language models operate through probabilistic pattern recognition, generating text based on statistical regularities in vast corpora of existing language. As a result, the outputs they produce tend toward aesthetic coherence, fluency, and familiarity, often relying on widely recognisable rhetorical structures and stylistic conventions (Bender et al., 2021). While this enables efficiency and accessibility, it also raises concerns about the standardisation and homogenisation of discourse, as well as the potential erosion of originality and critical depth (Floridi & Chiriatti, 2020).

In organisational settings, this may lead to the proliferation of texts that are persuasive in form but limited in substantive insight, which can be interpreted as a new, technologically mediated form of kitsch (Szostak, 2025). Also, critical aesthetic discourse on the role of AI in managerial and humanistic settings is still limited but growing (Szostak, 2026b, 2026a, 2026c).

This article analyses the concept of *algorithmic kitsch*, capturing its emerging phenomenon. Algorithmic kitsch refers to AI-generated results (writing in our case) that simulate meaning through aesthetic and rhetorical optimisation while remaining detached from authentic experience, intentionality, and reflexive understanding. By synthesising insights from aesthetic theory and management studies, with AI implementation, the paper aims to provide a conceptual framework for analysing the aesthetic and epistemic implications of AI-generated managerial written discourse as a crucial component of managerial communication. In doing so, it contributes to debates on authorship, authenticity, and the role of technology in shaping contemporary cultures of writing (Putri et al., 2025).

Method

This article adopts a conceptual and theory-building approach, situated within interpretive management research and informed by aesthetic philosophy. Its primary aim is to develop and refine the concept of *algorithmic kitsch* as a lens for understanding AI-generated managerial writing. The methodology integrates three strands of analysis: (1) a theoretical synthesis of literature on kitsch, management aesthetics, and AI-generated discourse (key monographs in the field and articles in English from databases: Google Scholar, Mendeley, Scopus, Web of Science); (2) conceptual abstraction, through which core attributes and mechanisms of algorithmic kitsch are identified; and (3) the construction of an analytical model that enables systematic comparison between different modes of writing (from sole human to sole AI-generated).

The analysis is guided by three research questions: (RQ1) How can AI-generated managerial and organisational written communication be

conceptualised through the phenomenon of *kitsch*? (RQ2) What aesthetic mechanisms underpin algorithmic kitsch in managerial and organisational communication? (RQ3) What are the epistemic implications of adopting AI--generated written communication in managerial practice?

Theoretical framework

Aesthetic insights into kitsch, management, and communication

The concept of *kitsch* has long occupied a central place in aesthetic theory as a category describing cultural artefacts that privilege accessibility, emotional immediacy, and formal familiarity over originality and critical depth. Classical accounts emphasise that kitsch operates through *simplification*, reducing complex realities to easily consumable forms, while offering *immediate emotional gratification* without requiring interpretive effort (Greenberg, 1985; Scruton, 2011). Its reproducibility (the capacity to replicate established aesthetic patterns) further reinforces its appeal, yet simultaneously contributes to a *lack of critical depth*, as meaning is pre-packaged rather than discovered. In this sense, kitsch does not merely imitate art but simulates its effects, producing a semblance of significance detached from authentic experience (Ryynänen & Barragán, 2023).

These characteristics are highly relevant in the context of *management aesthetics*, which conceptualises organisations as inherently *aesthetic systems* shaped by symbols, narratives, and sensory forms (Strati, 1999; Szostak, 2024a). Organisational communication (ranging from mission statements to strategic reports) can be understood as an *aesthetic artefact*, crafted to convey coherence, legitimacy, and emotional resonance (Barry & Elmes, 1997). Moreover, managerial language functions performatively, not only describing reality but actively constructing it through rhetorical and symbolic means (Boje, 2001). Within such contexts, kitsch emerges when communication becomes overly formulaic, emotionally stylised, and detached from substantive organisational practices (Hertzberg, 2017), resulting in what may be termed *aestheticised managerialism*.

The rise of AI-generated discourse intensifies these tendencies. Large language models operate through pattern-based generation, producing text by identifying and recombining statistical regularities in vast datasets (Bender et al., 2021). Consequently, their outputs prioritise *statistical plausibility* over *truth*, generating statements that are likely to be accepted as coherent rather than grounded in situated knowledge (Floridi & Chiriatti, 2020). At the same time, these systems are optimised for *readability* and *persuasion*, favouring clarity, fluency, and rhetorical balance. While such qualities enhance communicative efficiency, they also align closely with the defining features of kitsch: smoothness, familiarity, and emotional accessibility without epistemic depth (Szostak, 2025).

Taken together, these perspectives suggest that contemporary managerial practices are increasingly susceptible to forms of communication that can be described as kitsch-like, particularly when mediated by AI. The convergence of aesthetic simplification, performative language, and algorithmic generation creates conditions in which discourse may appear meaningful and authoritative while remaining largely simulated. This underscores the need for a critical, aesthetic analysis of managerial communication (including writing) in the age of AI.

AI-generated writing communication in managerial practice

The integration of generative AI into managerial communication practices is rapidly transforming how organisations produce, circulate, and interpret written texts. In contemporary organisations, AI-supported tools are increasingly used to generate a wide range of documents, including strategic reports, policy documents, internal emails, marketing content, ESG disclosures, performance evaluations, and stakeholder communications. These text types are not merely operational artefacts; they play a central role in shaping organisational identity, legitimising decisions, and constructing narratives of value and responsibility (Barry & Elmes, 1997; Christensen et al., 2013). The adoption of AI in producing such texts thus represents not only a technological shift but also a transformation in the cultural and symbolic dimensions of managerial practice.

A defining feature of AI-generated writing is its *fluency* and *coherence*. Large language models are designed to produce text that is grammatically correct, logically structured, and stylistically consistent, often exceeding human performance in terms of speed and surface-level clarity (Floridi & Chiriatti, 2020). This fluency contributes to the perception of professionalism and competence in organisational communication, particularly in contexts where clarity and standardisation are valued. Additionally, AI-generated texts typically exhibit a *neutral and balanced tone*, avoiding strong emotional expressions or controversial positions. This neutrality can be advantageous in formal communication settings, such as corporate reporting or stakeholder engagement, where consistency and diplomatic language are required.

However, these same properties also introduce significant risks. One of the most pressing concerns is the *over-standardisation of discourse*. Because AI systems generate text based on patterns learned from large datasets, they tend to reproduce dominant linguistic conventions and rhetorical structures. As a result, organisational communication may become increasingly homogenised, with different companies producing texts that are stylistically and conceptually similar. This phenomenon aligns with broader concerns about institutional isomorphism, where organisations adopt similar practices and narratives to gain legitimacy (DiMaggio & Powell, 1983). In the context of AI-generated writing, such isomorphism is not only institutional but also *algorithmic*, driven by shared training data and optimisation processes.

Moreover, the reliance on AI may lead to a subtle shift from *meaning-making* to *text production*. Writing, traditionally understood as a cognitive and reflective process, risks being reduced to the generation and editing of pre-structured outputs. Large language models function as "stochastic parrots", producing plausible word sequences without understanding their meaning (Bender et al., 2021). This raises concerns about the epistemic quality of managerial texts, particularly when AI-generated content is accepted uncritically as authoritative or insightful (Kingery-Page & Hahn, 2012). In ESG communication, for instance, the use of AI may amplify the production of rhetorically polished but substantively vague narratives, potentially reinforcing practices of symbolic compliance or *greenwashing* (Christensen et al., 2013).

At the same time, AI-generated writing reshapes managers' roles as communicators. Rather than crafting texts from scratch, managers increasingly act as editors, curators, and prompt designers, guiding AI systems to produce desired outputs. This shift has implications for authorship and responsibility, as the boundaries between human and machine contributions become blurred. It also places new demands on managerial competencies, including the ability to evaluate AI-generated content critically and to maintain coherence between organisational discourse and actual practices. While AI-generated writing offers clear benefits in terms of efficiency and consistency, it also introduces risks related to standardisation, epistemic shallowness, and the erosion of authentic organisational voice. These tensions highlight the need for a more critical and reflexive approach to the use of AI in managerial communication.

Results

The results are presented in the following order: (1) development of the algorithmic kitsch concept, (2) presentation of a selection of algorithmic kitsch mechanisms, (3) description of aesthetic effects of algorithmic kitsch, and (4) development of an analytic model of "AI-supported managerial written communication: From authentic expression to algorithmic kitsch".

Algorithmic kitsch concept

This section develops the concept of *algorithmic kitsch* as a theoretical construct capturing the aesthetic and epistemic qualities of AI-generated managerial discourse. Building on classical notions of kitsch and contemporary analyses of AI-generated text, *algorithmic kitsch* may be defined as "a mode of discourse produced by generative AI that achieves aesthetic coherence, rhetorical persuasiveness, and emotional accessibility through pattern-based generation, while lacking epistemic depth, originality, and authentic intentionality".

This definition positions algorithmic kitsch at the intersection of aesthetic theory and computational linguistics. Like traditional kitsch, it operates through

recognisable forms, simplified meanings, and emotional immediacy (Greenberg, 1985; Hoberman, 2005; Kulka, 1996). However, unlike earlier forms, it is generated not solely by human imitation but by algorithmic recombination of large-scale textual patterns derived from training datasets. In this sense, algorithmic kitsch represents a *technologically mediated intensification of pre-existing tendencies* in organisational communication.

Mechanisms of algorithmic kitsch

We can identify three underlying mechanisms that explain how algorithmic kitsch is produced in managerial writing contexts: (1) dataset-derived clichés, (2) optimisation for readability and persuasion, and (3) removal of friction and ambiguity.

Dataset-derived clichés. Generative AI systems rely on statistical regularities extracted from vast corpora of existing texts. Consequently, AI-generated texts tend to privilege high-frequency phrases, conventional structures, and culturally dominant rhetorical templates. In managerial communication, this results in the proliferation of familiar expressions (such as "driving innovation", "creating stakeholder value", or "committed to sustainability") which, while rhetorically effective, often lack specificity and originality. These dataset-derived clichés are not incidental but structurally embedded in the generative process, reinforcing the aesthetic logic of kitsch as repetition and imitation.

Optimisation for readability and persuasion. A second mechanism is the optimisation of outputs for clarity, fluency, and persuasive balance. Large language models are trained and fine-tuned to produce text that is easily interpretable and contextually appropriate, often avoiding ambiguity, contradiction, or stylistic irregularity (Floridi & Chiriatti, 2020). This results in discourse that is smooth, coherent, and rhetorically aligned with audience expectations, particularly in formal organisational contexts. However, this optimisation also reduces the presence of tension, uncertainty, and critical nuance, which are central to reflective thinking and meaningful argumentation. In this way, algorithmic kitsch transforms writing into a form of *aesthetic*

optimisation, in which communicative success is measured by plausibility and acceptability rather than by truth or insight.

Removal of friction and ambiguity. The third mechanism involves systematically removing cognitive and semantic friction. Human writing often reflects the complexity of thought, including hesitations, contradictions, and unresolved questions. In contrast, AI-generated text tends to eliminate such features, producing outputs that are internally consistent and free of ambiguity. While this enhances readability, it also flattens interpretive space, suppressing alternative meanings and critical perspectives. As a result, discourse becomes less dialogical and more monological, described as "aspirational talk"; language that signals intent without engaging with underlying complexities (Christensen et al., 2013).

Aesthetic effects of algorithmic kitsch

These mechanisms of algorithmic kitsch give rise to three interrelated aesthetic effects that define algorithmic kitsch in practice: (1) aesthetic smoothing (removal of tension), (2) emotional templating (predictable affect), and (3) semantic flattening (loss of depth).

Aesthetic smoothing (removal of tension). AI-generated texts exhibit a high degree of formal polish and internal coherence, often eliminating contradictions, dissonance, or stylistic irregularities. This *aesthetic smoothing* creates an impression of clarity and professionalism but reduces the presence of productive tension that is essential for critical reflection and innovation.

Emotional templating (predictable affect). Algorithmic kitsch relies on standardised emotional cues, producing texts that are affectively calibrated but emotionally predictable. This aligns with the logic of kitsch as immediate emotional gratification (Ortlieb & Carbon, 2019a, 2019b), yet in AI-generated discourse, such affect is not experienced but simulated through learned patterns.

Semantic flattening (loss of depth). Algorithmic kitsch leads to semantic flattening, where complex ideas are reduced to generalised, non-specific formulations. This diminishes the epistemic richness of discourse, replacing nuanced argumentation with broadly acceptable statements. As a result,

meaning is not explored but pre-structured and stabilised, limiting the potential for genuine knowledge creation.

Model: From authentic expression to algorithmic kitsch

The tool that can help identify the communication in the field of interaction between *authentic expression* and *algorithmic kitsch* is the analytical model presented in Table 1. The model is conceptualised as a continuum across five dimensions: (1) source of meaning (lived experience versus pattern replication), (2) voice (subjective versus simulated), (3) argumentation (reflective versus pre--structured), (4) aesthetics (expressive versus smooth and formulaic), and (5) ethics (responsibility-bearing versus responsibility-diffused). This model serves as a heuristic framework for interpreting how AI reshapes managerial discourse and for locating hybrid human-AI writing practices between authenticity and simulation.

Table 1. An analytical model of AI-supported managerial written communication: From authentic expression to algorithmic kitsch

Dimension	*Humanistic Writing*	*Hybrid (Human-AI)*	*Algorithmic Kitsch*
Source of meaning	Lived experience	Co-constructed	Pattern replication
Voice	Situated, subjective	Blended	Simulated
Argumentation	Reflective, effortful	Assisted	Pre-structured
Aesthetics	Expressive, sometimes imperfect	Polished	Smooth, formulaic
Ethics	Responsibility--bearing	Diffused	Displaced

Source: Author's elaboration.

Discussion

The concept of algorithmic kitsch, as developed in this study, has significant implications for managerial practice and organisational functioning. These implications unfold across epistemic, ethical, and aesthetic dimensions, each revealing tensions introduced by the integration of AI-generated writing into managerial communication.

Implications for managerial communication

Epistemic implications. One of the most profound consequences of algorithmic kitsch is the *illusion of knowledge production*. AI-generated texts often exhibit high levels of coherence, fluency, and rhetorical completeness, which may create the impression of well-developed reasoning and insight. However, such systems generate texts based on statistical associations rather than genuine understanding (Bender et al., 2021). This leads to a situation in which *plausibility* is mistaken for *knowledge*, and *surface-level coherence* substitutes for *epistemic depth*. In managerial contexts, this may result in decisions being supported by texts that appear analytically robust but are, in fact, epistemically shallow or derivative. Moreover, the thoughtless, unconscious reliance on AI-powered tools contributes to the replacement of *thinking* with *prompting*, where the cognitive effort traditionally associated with writing is displaced by the technical skill of generating and refining prompts. Writing, instead of being a process of *reflection* and *sensemaking*, risks becoming a process of *selecting* and *validating* pre-generated content, thereby weakening the epistemic foundations of managerial practice.

Ethical implications. Algorithmic kitsch also introduces complex ethical challenges, particularly concerning responsibility and authorship. As AI systems become co-producers of organisational discourse, responsibility for the content becomes increasingly diffused. Managers may rely on AI-generated outputs while simultaneously distancing themselves from their authorship, leading to ambiguity regarding accountability for decisions, claims, or representations embedded in the text. This raises the fundamental question:

"Who speaks?". If organisational texts are generated through algorithmic processes, the traditional notion of a responsible author is destabilised. This issue is particularly salient in areas such as ESG communication, where organisations are expected to demonstrate ethical commitment and transparency. When such narratives are partially or wholly AI-generated, there is a risk that ethical expression becomes *performative* rather than *substantive*, aligning with the notion of "aspirational talk" (Christensen et al., 2013). In this context, algorithmic kitsch may obscure the distinction between genuine commitment and symbolic representation.

Aesthetic implications. From an aesthetic perspective, the proliferation of AI--generated writing contributes to the dominance of *polished but empty* discourse. The mechanisms identified earlier (aesthetic smoothing, emotional templating, or semantic flattening) result in texts that are formally appealing yet lacking in tension, ambiguity, and depth. Such discourse aligns with the characteristics of kitsch as described in aesthetic theory: immediate, accessible, and emotionally resonant, yet ultimately superficial (Greenberg, 1985; Kulka, 1996).

At the organisational level, this may lead to the emergence of *simulacra*—representations that no longer refer to underlying realities but instead circulate as self-contained signs (Baudrillard, 1994). Managerial texts, particularly those generated or heavily supported by AI, risk becoming self-referential narratives of coherence and value, disconnected from actual practices and experiences. This aestheticisation of communication reinforces a form of organisational hyperreality, in which meaning is not constructed through *engagement with reality* but through the *reproduction of credible-looking discourse*.

Humanistic management perspective

In response to these challenges, a humanistic management perspective offers a critical corrective by reintroducing the centrality of dignity, authenticity, and responsibility in managerial practice (Pirson, 2017; Szostak, 2023). Humanistic management emphasises the role of the human actor as a moral and reflective agent, capable of meaning-making that goes beyond instrumental rationality and technical efficiency.

First, the principle of *dignity* calls for recognising the human subject not merely as a user of AI tools but as a bearer of intrinsic value and interpretive capacity. This implies that managerial communication should not be reduced to *optimised outputs* but should reflect *genuine engagement* with meaning, context, and consequence.

Second, *authenticity* becomes a critical aesthetic and ethical criterion. In contrast to algorithmic kitsch, which simulates expression through pattern replication, authentic managerial writing involves situated voice, intentionality, and experiential grounding (Nañola et al., 2025). Maintaining authenticity in AI-mediated contexts requires conscious effort to ensure that texts are not only coherent but also truthful representations of organisational realities.

Third, *responsibility* must be reaffirmed in the face of distributed authorship. Even when AI systems generate content, managers remain accountable for the messages conveyed and the decisions supported by such texts. This necessitates a shift from passive reliance on AI outputs to active critical evaluation and ethical stewardship.

Within this framework, the role of the human actor shifts from mere *editor* of AI-generated text to *curator of meaning*. This entails selecting, interpreting, contextualising, and, when necessary, resisting AI-generated suggestions in order to preserve the integrity of organisational discourse. The human manager thus becomes responsible not only for the content of communication but also for its aesthetic and ethical quality.

Conclusions

Answering RQ1 (How can AI-generated managerial and organisational written communication be conceptualised through the phenomenon of kitsch?), it can be said that AI-generated managerial and organisational written communication can be conceptualised as a form of kitsch insofar as it reproduces the defining characteristics of this aesthetic category: simplification, emotional immediacy, reproducibility, and lack of critical depth. However, the study demonstrates that AI does not merely replicate existing forms of kitsch but industrialises

and systematises them. Generative AI produces discourse that simulates meaning, coherence, and intentionality without grounding in lived experience or reflective cognition. Algorithmic kitsch thus represents a technologically mediated intensification of aestheticised managerial communication, where plausibility and fluency replace authenticity and epistemic engagement.

Answering RQ2 (What aesthetic mechanisms underpin algorithmic kitsch in managerial and organisational communication?), it can be stated that there are at least three mechanisms underpinning algorithmic kitsch in managerial communication: (a) dataset-derived clichés, (b) optimisation for readability and persuasion, and (c) the removal of friction and ambiguity. These mechanisms give rise to three corresponding aesthetic effects: (a) aesthetic smoothing, (b) emotional templating, and (c) semantic flattening. The produced texts are rhetorically effective yet structurally predictable and epistemically shallow. The analytical model developed in this study (from authentic expression to algorithmic kitsch) demonstrates how AI--mediated communication shifts managerial discourse toward simulation, standardisation, and aesthetic optimisation. This framework offers a tool for diagnosing the extent to which organisational communication retains authentic expressive qualities or is subsumed by algorithmic reproduction.

Answering RQ3 (What are the epistemic implications of adopting AI--generated written communication in managerial practice?), it can be said that the epistemic implications of adopting AI-generated writing in managerial practice are substantial. The findings highlight the emergence of an illusion of knowledge production, where fluency and coherence create the appearance of insight without substantive understanding. This is accompanied by a shift from thinking to prompting, in which the cognitive and reflective dimensions of writing are displaced by procedural interaction with AI systems. As a result, managerial communication risks losing its function as a site of sensemaking and becoming a mechanism of textual production and validation. This shift has profound consequences for decision-making, organisational learning, and the integrity of knowledge processes.

Despite its contributions, this study has several *limitations*: (1) it is conceptual in nature and does not include empirical validation of the proposed framework;

while the theoretical synthesis provides analytical clarity, future studies should test the concept of algorithmic kitsch through qualitative or quantitative analyses of real organisational texts; (2) the study focuses primarily on written communication, leaving aside other forms of AI-mediated expression (e.g., visual or multimodal content) that may exhibit similar dynamics; (3) the analysis is largely grounded in Western aesthetic and management traditions, which may limit its applicability across different cultural contexts; (4) the evolving nature of AI technologies means that the characteristics of algorithmic kitsch may shift as models become more advanced and context-aware.

These limitations open several avenues for *future research*: (1) empirical studies could investigate the presence of algorithmic kitsch in specific domains, such as ESG reporting, corporate communication, or academic writing, using discourse analysis or corpus linguistics methods; (2) comparative studies between human-authored, AI-assisted, and fully AI-generated texts could further refine the proposed continuum model; (3) another promising direction involves exploring the perception of algorithmic kitsch among stakeholders, including employees, managers, and external audiences, to assess its impact on trust, legitimacy, and organisational identity; (4) interdisciplinary research could examine how algorithmic kitsch interacts with broader phenomena such as digital capitalism, platformisation, and the commodification of language; (5) future research should also engage more deeply with the principles of humanistic management, particularly in developing frameworks for maintaining authenticity, dignity, and responsibility in AI-mediated communication, including exploring educational and organisational interventions that enhance aesthetic literacy and critical reflexivity, enabling managers to navigate the tensions between efficiency and meaning.

References

Barry, D., & Elmes, M. (1997). Strategy Retold: Toward a Narrative View of Strategic Discourse. *Academy of Management Review*, *22*(2), 429–452. DOI: 10.5465/amr.1997.9707154065.

Baudrillard, J. (1994). *Simulacra and Simulation*. University of Michigan Press.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). DOI: 10.1145/3442188.3445922.

Boje, D. (2001). *Narrative Methods for Organizational & Communication Research*. SAGE Publications, Ltd.

Christensen, L. T., Morsing, M., & Thyssen, O. (2013). CSR as aspirational talk. *Organization*, *20*(3), 372–393. DOI: 10.1177/1350508413478310.

DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, *48*(2), 147. DOI: 10.2307/2095101.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, *30*(4), 681–694. DOI: 10.1007/s11023--020-09548-1.

Greenberg, C. (1985). Avant-garde and Kitsch. In: F. Frascina (Ed.), *Pollock and After: The Critical Debate*. Harper and Row.

Hertzberg, J. (2017). AI, Kitsch, and Communication. *KI–Künstliche Intelligenz*, *31*(4), 313–315. DOI: 10.1007/s13218-017-0508-6.

Hoberman, R. (2005). Aesthetic Taste, Kitsch, and the Years. In: *Woolf Studies Annual* (Vol. 11, Issue 77, pp. 77–98). Pace University Press.

Kingery-Page, K., & Hahn, H. (2012). The aesthetics of digital representation: realism, abstraction and kitsch. *Journal of Landscape Architecture*, *7*(2), 68–75. DOI: 10.1080/18626033.2012.746091.

Kostera, M. (1997). The Kitsch-Organisation. *Studies in Cultures, Organizations, and Societies*, *3*(2), 163–177. DOI: 10.1080/1024528970852349 4.

Kulka, T. (1996). *Kitsch and Art*. Pennsylvania State University Press.

Linstead, S. (2002). Organizational kitsch. The Aesthetic Turn in Management. *Organization*, *9*(4), 657–682. DOI: 10.4324/9781351147965-4.

Linstead, S., & Höpfl, H. J. (2000). *The Aesthetics of Organization*. SAGE Publications.

Nañola, E. L., Arroyo, R. G. L., Hermosura, N. J. T., Ragil, M., Sabanal, J. N. U., & Mendoza, H. B. (2025). Recognizing the artificial: A comparative voice analysis of AI-Generated and L2 undergraduate student-authored academic essays. *System*, *130*, 103611. DOI: 10.1016/j.system.2025.103611.

Ortlieb, S. A., & Carbon, C. C. (2019a). A functional model of kitsch and art: Linking aesthetic appreciation to the dynamics of social motivation. *Frontiers in Psychology*, *9*(2437), 1–17. DOI: 10.3389/fpsyg.2018.02437.

Ortlieb, S. A., & Carbon, C. C. (2019b). Kitsch and perception: Towards a new "aesthetic from below". *Art and Perception*, *7*(1), 1–26. DOI: 10.1163/22134913-00001091.

Pirson, M. (2017). *Humanistic Management. Protecting Dignity and Promoting Well-being*. Cambridge University Press.

Putri, H., Susilo, D., Munandar, E., Yasin, H., & Atmaja, I. (2025). The Algorithmic Gaze: Deconstructing Authorship and Aesthetics in Generative Artificial Intelligence (AI) Art. *Enigma in Cultural*, *3*(1), 55–66. DOI: 10.61996/cultural.v3i1.106.

Ryynänen, M., & Barragán, P. (Eds.) (2023). *The Changing Meaning of Kitsch. From Rejection to Acceptance*. Palgrave Macmillan Cham.

Scruton, R. (2011). *Beauty: A Very Short Introduction*. Oxford University Press.

Strati, A. (1999). *Organization and Aesthetics*. SAGE Publications Ltd.

Szostak, M. (2021). Artistry, Management, and Creativity: Links and Common Denominators. *Discourses on Culture*, *16*(1), 23–54. DOI: 10.36145/DoC2021.08.

Szostak, M. (2023). Reflexivity in aesthetic situation management. *Discourses on Culture*, *20*, 161–197. DOI: 10.2478/doc-2023-0015.

Szostak, M. (2024a). *Humanistic Management, Organisation and Aesthetics: Art of Management and Management of Art*. Routledge. DOI: 10.4324/9781003458029.

Szostak, M. (Ed.) (2024b). *Non-Artistic Kitsch*. Palgrave Macmillan. DOI: 10.1007/978-3-031-72237-0.

Szostak, M. (Ed.) (2025). *Management Aesthetics. Kitsch and Modern Organisations*. Routledge. DOI: 10.4324/9781003504931.

Szostak, M. (Ed.) (2026a). *AI and Humanistic Management: Aesthetics in Managerial Theory and Practice*. Routledge. DOI: 10.4324/9781003654377.

Szostak, M. (Ed.) (2026b). *Organizations, Humanism and AI: An Aesthetics Approach*. Routledge. DOI: 10.4324/9781003662723.

Szostak, M. (Ed.) (2026c). *Aesthetics of Human-AI Collaboration in Creative Activities: Art, Cultural Heritage, and Academic Research Practice*. Brill | mentis. DOI: 10.30965/9783969753460.

DOI: 10.2478/doc-2026-0003

This is an open access article licensed under the Creative Commons AttributionNonCommercial-NoDerivatives 4.0 International (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Rania Za'rour

Department of English Language and Literature, The University of Jordan, Amman, Jordan

r.zarour@ju.edu.jo

ORCID ID: 0000-0001-9873-6099

Abdel Rahman Mitib Altakhaineh

Department of English Language and Literature, The University of Jordan, Amman, Jordan

a.altakhaineh@ju.edu.jo

ORCID ID: 0000-0001-7605-2497

False Positive Risk in AI Detection of L2 Academic Writing: A Case Study

Article history:

Received **16 March 2026**

Revised **25 April 2026**

Accepted **26 April 2026**

Available online **23 June 2026**

Abstract: AI-text detectors are increasingly used in education and scholarly publishing. While these tools are meant to protect academic integrity, they may also create new inequities for multilingual writers. This exploratory case study investigates why legitimate second language (L2) academic prose can be misclassified as AI-generated. We start from the premise that L2 academic writing functions as a distinct register whose recurrent linguistic features create predictability, which is one of the main metrics to detect AI-generated text. Using a single-author corpus repeatedly flagged as machine--generated, we develop a detector-inspired simulation capturing five proxies of high predictability: low perplexity, low burstiness, high formulaic language, consistent formality, and low lexical diversity. We simulated two common detection models: a prevalence model (flagging frequent proxies) and a cluster--based model (flagging co-occurrence of features). Prevalence-based detection flagged 62% of the sample manuscripts, while cluster-based detection flagged 38% with three or more co-occurring proxies. Although dense clustering (four to five proxies) was uncommon (7.8%), our sample was repeatedly flagged by both logics. These findings do not establish equivalence with commercial detection systems, but they suggest that highly conventionalized L2 academic prose may face a heightened risk of false positive AI flagging. We term this phenomenon the "detection paradox".

Keywords: AI detection, academic stratification, L2 writing, multilingual authors, linguistic bias

Introduction

AI-generated text detection is increasingly used in editorial workflows, academic integrity procedures, and writing assessment, shaping desk decisions and author feedback (Kehkashan et al., 2025). AI-generated text detection refers to automated procedures that estimate the likelihood that a text was produced by a large language model rather than a human. Detector outputs are, nevertheless, probabilistic and sensitive to genre, writer background, and text preprocessing (Liang et al., 2023; Weber-Wulff et al., 2023; Wang et al., 2024). The operational adoption of detection tools grants them practical authority, even as concerns about their generalizability and fairness persist in the literature (Abd-Elaal et al., 2022; Giray et al., 2025; Kehkashan et al., 2025). Because institutional workflows amplify the consequences of automated signals, false positives carry higher operational risk (Giray, 2024), particularly for second-language English authors, whose legitimate stylistic patterns can be mistaken for the statistical regularities that detectors interpret as machine-generated (Giray et al., 2025).

While AI detection-related false positives may affect a range of writer populations, this study focuses specifically on second language (L2) academic writers. Prior research suggests that L2 academic writing, even at advanced proficiency levels and across diverse L1 backgrounds, often exhibits discourse and stylistic patterns that distinguish it from prototypical L1 academic prose. These include greater reliance on formulaic sequences (Paquot & Granger, 2012), more explicit cohesive devices (Hinkel, 2001), more conservative lexical choices (Laufer & Waldman, 2011), more repetitive sentence structures (Hinkel, 2003; Lu, 2011), and closer adherence to genre conventions (Hyland, 2016). At the same time, we do not claim that these features are unique to L2 writers; highly conventionalized L1 academic prose may display similar configurations (Shin, 2019). Our argument, rather, is that the probabilistic clustering of such features may be especially common in some forms of L2 academic writing, thereby increasing the likelihood that multilingual authors are misread by current detection heuristics despite producing legitimate human-authored texts.

Critically, these L2 features closely resemble statistical patterns that many AI detection approaches are designed to capture (Opara, 2025). Most detectors

operationalize human writing as relatively variable and AI writing as relatively uniform, using measures tied to predictability, sentence- and paragraph-level variability, and repetition (Gehrmann et al., 2019; Giray, 2024; Kar et al., 2025; Liu et al., 2024; Mitchell et al., 2023; Picazo-Sanchez & Ortiz-Martin, 2024; Showemimo, 2025). Pattern predictability is typically operationalized using two models of detection. The first identifies localized patterns, flagging text when a threshold of co-occurring heuristics is surpassed within a document. The second analyzes document-wide heuristics to generate aggregate percentage scores representing the estimated proportion of AI-generated content.

Nevertheless, these statistical cues are not unique to AI-generated text (Doughman et al., 2025). Academic L2 writing's formulaic expressions and repetitive structures increase local predictability and lower perplexity. Additionally, conservative lexical choices reduce lexical dispersion, while adherence to genre conventions produces uniform syntactic patterns, all of which are treated as hallmarks of AI-generated prose (Muñoz-Ortiz et al., 2024). We argue that these cues are more likely to co-occur in legitimate L2 academic writing, especially in highly conventional sections, which makes multilingual writers structurally more vulnerable to false positives, with predictable downstream effects for educational assessment and publication gatekeeping (Ahmed & Ahmed, 2026; Giray et al., 2025; Kar et al., 2024; Liang et al., 2023).

Empirical results underline the practical consequence of this miscalibration. In some evaluations, more than 60% of NNS essays were incorrectly labeled as AI-generated by widely used detectors, indicating that false positives are not a marginal edge case but a systematic risk for multilingual writers (Giray, 2024; Kirchenbauer et al., 2023; Liang et al., 2023). Further evidence suggests that this risk intensifies when multiple L2 features co-occur, particularly in specialized academic genres where disciplinary conventions further constrain linguistic variation (Flitcroft et al., 2024; Giray et al., 2025; Popkov & Barrett, 2025).

This study is motivated by our experience with the heightened risk of AI detection as L2 academic writers. Despite high proficiency and a strong publication record, multiple manuscripts by the first author were flagged during editorial screening, constituting a recurring pattern and an empirically consequential point of departure for the study. To investigate this pattern in

a controlled and interpretable manner, we frame the present analysis explicitly as a single-author case study. The corpus was intentionally limited to one proficient L2 academic writer for two primary reasons. First, the author used AI tools in a controlled manner while writing the sample manuscripts, opting for minimal grammatical checking. Second, this approach aimed to reduce inter--author stylistic variation, allowing for a thorough examination of the textual features that could have triggered AI flagging in this specific case. Accordingly, the findings are exploratory and analytical rather than broadly generalizable, and they are not intended to support sweeping claims about L2 writers as a whole. Building on evidence that detectors often rely on predictability-oriented statistics (Miralles-González et al., 2025) and recognizing that commercial detectors are proprietary and opaque (Tufts et al., 2025), often providing only aggregate AI percentages, we engineered a transparent AI-detection simulator that operationalizes these heuristics under each hypothesis. This approach is designed to test plausible triggering mechanisms rather than to claim empirical equivalence with commercial systems. We do not benchmark proprietary commercial tools because their opacity precludes inference about why texts are flagged and limits comparison to surface-level similarity in outcomes rather than explaining underlying causes. Since our goal is to investigate mechanisms of false positive risk rather than reproduce black-box outputs, the simulator provides a more interpretable and reproducible basis for analysis. Specifically, we address the following research questions:

1) To what extent do these proxies, individually and in combination, yield elevated suspicion scores within our L2 corpus?
2) How are these simulated suspicion scores distributed across research article sections (e.g., Introduction, Methods, Results, Discussion)?
3) Which proxy configurations characterize the most highly suspicious segments, and how do these configurations correspond to known features of conventionalized academic prose?

Literature review

L2 academic writing

Empirical evidence demonstrates that L2 academic writing, particularly in English, exhibits characteristic linguistic patterns that distinguish it from L1 academic prose (Subandowo et al., 2025), even at advanced proficiency levels (Zaini & Ollerhead, 2019). These features arise not from deficient or developing L2 competence, but from legitimate L2 writing strategies, cross-linguistic influence (Jarvis & Pavlenko, 2008), and cognitive resource management during L2 composition (Zaini & Ollerhead, 2019). L2 writers strategically draw on these patterns to meet academic genre conventions, achieve clarity and precision in a non-native language, and produce manuscripts that satisfy the linguistic expectations of global scholarly publication and evaluation systems (Flowerdew, 2001; Hyland, 2016). For non-Western, non-Anglophone scholars, mastery of such conventionalized patterns is essential for participation in English-dominant academic discourse communities (Canagarajah, 2002; Lillis & Curry, 2010). Recognizing these features is crucial for understanding how L2 texts may interact with automated detection systems sensitive to lexical repetition and reduced diversity (Mitchell et al., 2023; Gehrmann et al., 2019).

For example, L2 academic writing is marked by heavy use of formulaic sequences and lexical bundles, that is, recurrent multi-word units in academic discourse (Paquot & Granger, 2012). Compared to L1 prose, L2 texts rely on more repetitive scaffolds and a narrower range of lexical combinations (Pan, 2018; Taşçı & Öztürk, 2021), which support cohesion, signal rhetorical moves, and maintain disciplinary register (Kashiha & Chan, 2015). This reliance is a strategic adaptation, as L2 writers draw on high-frequency academic phrases to reduce cognitive load and focus on content (Ädel & Erman, 2012). In technical sections, L2 academic writers further restrict lexis to preserve terminological precision, a tendency especially visible in STEM disciplines (Haq et al., 2023).

More broadly, L2 academic prose often shows simplified lexis and conservative word choice (Laufer & Waldman, 2011), with reduced lexical diversity, preference for high-frequency academic vocabulary, and limited use of low-frequency near

synonyms. Reusing familiar items promotes clarity and minimizes risks of register violations (Jarvis & McNamara, 2012), while repeated exposure to high-utility formulaic sequences and stable collocational patterns, rather than native-speaker stylistic norms, drives fluency in academic language (Casal & Yoon, 2023; Du & Hashimoto, 2023). At the syntactic level, L2 writers similarly lean on recurrent sentence templates and familiar clause-combining routines (Hinkel, 2003; Lu, 2011), recycling entrenched patterns to meet processing demands and maintain accuracy in high-stakes genres (Al Fattah, 2018; Canli & Yağiz, 2024).

Relatedly, information structure and thematic organization in L2 academic prose are strongly shaped by cross-linguistic influence. Because information packaging is language-specific, non-native English-speaking academics often transfer L1 discourse preferences into their L2 writing (Halliday & Matthiessen, 2014; Lambrecht, 1994), producing L1-patterned topicalization, theme placement, and clause-initial adverbials that affect how information is introduced and sequenced (Hinkel, 2002; Shehata & Eldakar, 2018). These patterns align with known mechanisms of second language acquisition and cross-linguistic influence (Murdoch et al., 2021). For example, L1-based thematic progression routines and preferences for clause-initial topics and discourse-organizing adverbials persist in L2 English (Pedersen, 2011; Mu & Zhang, 2018). When such L1-driven routines interact with academic genre norms that already favour predictable move structures and cohesive patterns, the resulting prose can appear highly patterned and formulaic, especially for writers with still-developing grammatical resources (Connor, 1996; Mu & Zhang, 2018). Cohesion is further shaped by a preference for overt discourse signaling, as L2 writers use more explicit cohesive devices like conjunctive adverbials, demonstratives, and lexical repetition, compared with L1 writers (Hinkel, 2001). This pattern reflects pedagogical emphasis and strategic compensation for perceived limits in implicit cohesion.

Patterns of stance and authorial presence also distinguish L2 texts. Non-native authors often underuse or redistribute hedges and self-references, reflecting cross-cultural rhetorical norms rather than simple linguistic deficiency (Ajideh et al., 2025; Ganjavi et al., 2024). These differences are especially salient in argumentative and evaluative sections, producing profiles of epistemic caution, engagement, and author visibility that may diverge from dominant

Anglophone conventions while remaining functionally effective (Mu & Zhang, 2018; Shchemeleva, 2021). Collectively, these information-structuring, cohesive, and stance-taking practices reinforce highly regular and explicitly signposted patterns of discourse organization in L2 academic writing.

The non-native academic writing genre itself exhibits disciplinary and national heterogeneity, with error profiles, bundle usage, and rhetorical patterns clustering by L1 and cultural proximity (Albelihi & Al-Ahdal, 2024). For example, East Asian L2 writers demonstrate patterns distinct from Romance-language or Arabic L2 writers in terms of syntactic complexity, lexical diversity, and rhetorical organization (Mavrou & Chao, 2023; Altakhaineh et al., 2024; Eid & Mutahar, 2025). These L1-specific patterns reflect both typological differences between languages and culturally mediated rhetorical traditions.

Disciplinary context also shapes L2 writing patterns. Different sections of academic research articles follow well-established stylistic conventions, with Methods and Results sections being typically formulaic and terminologically consistent (Ellis et al., 2008). This patterned variation means that highly standardized sections naturally exhibit controlled phrasing and reduced linguistic variance, making them comparatively easier for L2 authors to write, especially in STEM disciplines where shared technical conventions reduce language variance (Haq et al., 2023). In contrast, the Introduction and Discussion sections remain the most challenging writing tasks for L2 writers, as these sections demand more rhetorically complex, persuasive, and culturally mediated writing (Valipour et al., 2017; Bajwa et al., 2020; Junina et al., 2025). Argumentative sections require nuanced stance-taking and authorial voice that are especially challenging for L2 writers (Mu & Zhang, 2018; Shchemeleva, 2021).

L2 academic writing is therefore far from monolithic. Error profiles, bundle usage, and rhetorical routines vary systematically by L1 background, disciplinary field, and article section. Yet across this heterogeneity, tendencies toward regularity, repetition, and constrained variation at multiple levels of language persist. These features, we propose, closely resemble the surface patterns that contemporary AI detection tools target through proxies such as lexical diversity, burstiness, and formulaicity, providing the rationale for the next subsection's focus on AI detection methods and mechanisms.

AI detection technologies: Methods and mechanisms

AI-generated text detection has evolved rapidly since the widespread deployment of large language models, developing along partially overlapping chronological and thematic lines (Kehkashan et al., 2025). At their core, these detection tools operate by identifying predictable linguistic patterns, whether at the word level, sentence level, inter-sentence level, or across broader stylistic and genre conventions. We hypothesize that this fundamental mechanism creates a critical problem. Post-hoc detectors cannot reliably distinguish between (1) AI-generated text, which exhibits statistically probable patterns by design, and (2) L2 academic writing, which employs strategically formulaic patterns as a compensatory literacy strategy.

AI detection tools employ two primary methodological approaches, both of which rely on pattern predictability. Clustering-based models identify localized patterns characteristic of AI generation, flagging text when a threshold number of such features co-occur within a document. These clusters typically include high lexical diversity within constrained semantic fields, consistent sentence-level complexity, uniform transition patterns between ideas, and standardized academic register markers. Tools like GPTZero, Originality.AI, and Writer.com exemplify this approach. Prevalence-based models, by contrast, analyze document-wide statistical features to generate aggregate percentage scores representing the estimated proportion of AI-generated content. These models calculate frequencies of occurrence across target texts, measuring rates of formulaic expressions, transitional phrases, hedging devices, and syntactic templates, then produce normalized statistical values. Tools like Turnitin, Copyleaks, and Winston AI employ prevalence-based or hybrid detection mechanisms that combine both clustering and frequency analysis. Nevertheless, commercial vendors of AI detection tools uniformly fail to disclose their detection thresholds, i.e., the specific point at which their algorithms categorize text as *probably* machine-generated versus human-authored. This opacity applies across widely-used tools and prevents independent validation, also obscuring the empirical basis for classification decisions (Kehkashan et al., 2025). Understanding the technical foundations of these methods is essential

for evaluating their performance characteristics, limitations, and potential sources of systematic bias.

Many early detectors analyze token-level probabilities. A language model assigns a probability to each word given its context, and these signals are aggregated to distinguish human from machine-generated text (Gehrmann et al., 2019). Three related heuristics are central: perplexity, burstiness, and token-level likelihood. Perplexity captures how predictable a sequence is overall; lower perplexity indicates more predictable text. Because large language models favour high-probability continuations, AI-generated text tends to have lower perplexity than human writing (Flitcroft et al., 2024). Burstiness indexes variation in predictability across sentences. Human texts often alternate between more and less predictable passages, while AI outputs tend to be more uniform and thus less bursty (Popkov & Barrett, 2025). Token--probability heuristics apply thresholds to per-token likelihoods and their dispersion, flagging texts that consistently select high-probability tokens with low variance. These metrics are frequently combined into a composite detector score (Flitcroft et al., 2024; Popkov & Barrett, 2025). For L2 academic writing, these proxies align closely with the features described in the literature review. Legitimate L2 strategies can therefore depress perplexity, reduce burstiness, and concentrate probability mass on high-frequency tokens in ways that superficially resemble AI generation. Although statistical detectors can perform well under constrained conditions (same-domain text, sufficient length, minimally edited outputs), they are sensitive to domain shift, short inputs, and post-hoc editing (Gehrmann et al., 2019; Flitcroft et al., 2024), creating scope for systematic misclassification of patterned but human-authored L2 prose.

A second family of tools uses discriminative classifiers rather than explicit probability thresholds. In supervised settings, models are trained on labelled human and AI-generated texts and learn to separate them using features at multiple linguistic levels (Guo et al., 2023). In inference, an input is encoded and passed through the classifier, which outputs a probability or label indicating whether it more closely resembles human or AI examples; when training and test distributions align, such systems can achieve strong in-domain accuracy (Uchendu et al., 2020).

Zero-shot and representation-based detectors extend this idea by leveraging pretrained language models (e.g., BERT, RoBERTa) without task-specific fine--tuning. Instead of learning from large, labelled AI versus human corpora, they use embeddings or attention patterns as general-purpose representations of linguistic structure and infer whether a text is more "machine-like" or "human--like" (Ippolito et al., 2020). These approaches reduce the need for dedicated training data and, in principle, can detect outputs from novel generative systems without retraining. However, both supervised and zero-shot detectors inherit constraints from their data and architectures. Performance degrades under domain shift, when texts differ in genre, topic, register, or language variety (Aldeen et al., 2025), and models can often be evaded through paraphrasing or synonym substitution that preserves meaning while altering surface features (Guo et al., 2025). Zero-shot systems, in particular, inherit the biases and discourse norms of their pretrained models, which are largely based on English Wikipedia, academic publications, and web text reflecting Western academic and journalistic conventions (Bender et al., 2021). This creates a specific risk for L2 academic writing. The lexical, syntactic, and discourse-level regularities can become precisely the stylometric and representational features that classifiers associate with "machine-like" text.

Training data and benchmark limitations

The effectiveness of AI detection systems is limited by the datasets used for their development. These datasets largely focus on English and traditional academic styles, reflecting the demographics of major AI research communities (Wang et al., 2024). Consequently, detectors are trained on data dominated by Western, English L1 writers (Gosselin, 2025) and fail to adequately represent World Englishes, L2 academic writing, and non-Western rhetorical styles (Caines et al., 2023; Won et al., 2025). These systems often equate "human" text with Anglophone native-speaker norms, overlooking legitimate L2 strategies as "machine-like" (Liang et al., 2023; Weber-Wulff et al., 2023). Features common in L2 academic writing may be misinterpreted as negative signals in AI training, raising concerns about detector validity for diverse global academic texts.

Current systems focus more on surface-level statistical patterns than on deep semantic markers of authorship. Evidence shows that even minor paraphrasing can significantly lower detector scores, suggesting that classification depends on distributional cues rather than a clear notion of "authorship authenticity" (Liu et al., 2024).

Methodology

Study design

This study uses an experimental model-building approach to reverse-engineer AI detection mechanisms and test how they behave on a small sample of authentic L2 academic writing. The design responds to a central limitation in current practice. Commercial vendors and editorial offices provide no transparency about how their tools classify text as "AI-generated" or which features trigger flags. Without access to proprietary algorithms or detailed feedback on individual decisions, we cannot directly observe why specific passages in our corpus were misclassified. We therefore adopt a theory-driven simulation of the two dominant AI detection paradigms, i.e., clustering models and prevalence models, and ask whether the L2 writing patterns in our data systematically align with the heuristics these paradigms operationalize. Rather than attempting to replicate any single commercial system, we approximate the core proxy features that detectors are known to target and examine how often legitimate L2 prose exhibits those signatures.

To do this, we analyzed a single-author corpus to control for inter-author stylistic variation while maintaining ecological validity. The corpus comprised four research articles (approximately 28,000 words) written by the first author between 2022 and 2024. All four were rejected by peer-reviewed journals on suspicion of AI generation and flagged by commercial detection tools, despite only minimal use of generative AI (restricted to grammatical checking). Two articles address applied statistics topics and two address applied linguistics topics, each following the standard IMRD structure. These articles are used

strictly as data and are not cited as references in this manuscript to preserve anonymity during review.

This design offers several methodological advantages. First, it is controlled for AI-aided writing. Moreover, single authorship helps isolate L2-typical features from idiosyncratic personal style. Using real rejection cases, rather than artificially constructed texts, grounds the analysis in actual editorial outcomes and provides ecological evidence for the detection paradox. The inclusion of both technical and more discursive articles allows us to explore whether detection patterns are section- and discipline-specific or generalize across domains. The corpus yields 334 analyzable segments, providing sufficient granularity for feature-level statistical analysis. Finally, because the same author and texts were previously flagged by commercial tools, we can assess whether our reverse-engineered detector reproduces similar flagging patterns, offering a construct-valid approximation of how current detection architectures respond to L2 academic writing.

The AI detection simulator

Rather than running black-box commercial tools directly, we implement an AI detection simulator grounded in the literature on detection mechanisms and L2 writing characteristics. This simulator is a proof-of-concept tool; our aim is diagnostic, i.e., to test whether L2 writing strategies produce the specific statistical signatures that contemporary detectors interpret as evidence of machine generation. Instead of first coding the corpus for generic L2 features (e.g., formulaic sequences, transition markers, hedging devices) and then passing it through opaque detectors that return only aggregate scores, we code the corpus directly for AI detection heuristics that correspond to those L2 patterns. This allows us to examine, at the level of individual segments, whether the same proxies associated with AI output, i.e., high predictability, low variability, formulaicity, constrained lexis, are also systematically present in authentic L2 academic prose. The five features used in the simulator are deliberately simplified proxies that attempt to capture broad tendencies rather than precise linguistic or cognitive constructs.

The simulator operationalizes two competing detection logics. Clustering, where segments are flagged when multiple heuristics co-occur, and prevalence, where segments are flagged when individual heuristics exceed a threshold. In both cases, we focus on five proxy features identified in the literature as characteristic of both AI-generated text and L2 academic writing:

1) Low perplexity: predictable word sequences, measured via n-gram probabilities.
2) Low burstiness: uniform sentence-length distribution, captured by the coefficient of variation.
3) High formulaic language: reliance on recurrent multi-word expressions, measured by n-gram frequency.
4) Consistent formality: minimal variation in register, operationalized as low variance in formality scores.
5) Low lexical diversity: restricted vocabulary range, measured via type–token ratio.

All five features are implemented using computational measures rather than manual coding (see Appendix A for technical details). This choice has three main motivations. First, computational metrics provide objective, replicable quantification of linguistic patterns, reducing subjective coder bias. Second, they enable efficient analysis of a corpus size that would be impractical to annotate manually. Third, and most importantly, they simulate the statistical signatures that AI detection algorithms themselves target: we use comparable language models (e.g., GPT-2) and metrics (e.g., perplexity, burstiness) to approximate how detectors "see" the text.

Window-based analysis framework

To analyze how AI-detection-prone features are distributed across the papers, we used a window-based analysis. Instead of treating each article or section as a single unit, window-based analysis divides the text into many smaller, overlapping pieces (windows) of fixed length. This makes it possible to

see where particular patterns cluster in the text, rather than assuming that linguistic features are uniform across an entire article. In this study, each article was split into overlapping 200-word windows. Each 200-word window was treated as an independent unit of analysis. For every window, we calculated the five AI-detection proxies described above and noted whether each proxy crossed its threshold. We then assigned each window a suspicious score from 0 to 5, defined as the number of proxies that were flagged in that window. A score of 0 means that none of the AI-like features were present; a score of 5 means that all five features occurred together in the same 200-word span.

A window was classified as suspicious when it showed two or more flags. The two-flag threshold was chosen because a single flagged feature can simply reflect an isolated stylistic choice, whereas the co-occurrence of multiple features is closer to the pattern that AI detectors are designed to identify. Importantly, the term *suspicious* does not imply wrongdoing or that the text is actually AI-generated. It simply indicates that the window displays the statistical profile that detection tools are likely to interpret as machine-like. We report results at three levels of suspicion:

- 2+ flags (moderate suspicion): at least two of the five features present; this is the baseline threshold at which an AI detector might begin to treat a passage as questionable.
- 3+ flags (high suspicion): at least three features present; this indicates a stronger convergence of AI-like patterns.
- 4+ flags (very high suspicion): four or all five features present; this represents an extreme concentration of AI-like signals that is relatively rare in typical human writing but can still occur in highly conventionalized L2 academic prose.

For each suspicious window, we also recorded which specific features were flagged so that we could analyze common combinations (for example, whether low perplexity tends to co-occur with high formulaicity). These co-occurrence patterns are reported in the Results section. To illustrate how the flagging system

works, consider Window 35, an excerpt from the discussion section of one of the linguistics papers ("Linguistics 2"):

> "generalizable finding in the data. Positive face (PF) strategies, which focus on building solidarity and positive social relations (like appreciation, gratitude, motivation, and starting conversations), also show a positive effect on acceptability (+0.37). However, the p-value of 0.081 means this effect is not statistically significant at the conventional 0.05 level…"

This passage reads like typical discussion-section prose: it reports statistical results in formal academic language and then interprets them. In our simulator, this window received a suspicious score of 4 out of 5. It was flagged as:

- Low perplexity (F1): the wording is highly predictable because it uses standard phrases such as "positive effect on acceptability" and "statistically significant at the conventional 0.05 level".
- Low burstiness (F2): sentence structure and length are relatively uniform, producing a smooth, regular rhythm.
- High formulaic language (F3): the passage relies on common academic word combinations and technical expressions.
- Low lexical diversity (F5): the vocabulary is somewhat restricted, with repeated use of key terms related to statistical reporting.

By contrast, formality variance (F4) was not flagged in this window because the text moves slightly between technical reporting and more explanatory language, creating some register variation. This is genre-appropriate for interpretive social science writing and shows that even windows with high suspicious scores can still contain subtle stylistic shifts. (Full numeric details for this and additional examples are provided in Appendix B). This window-based setup directly mirrors the two detection logics modelled in our simulator. At the prevalence level, we ask how often individual proxies cross their thresholds across all windows, whereas at the clustering level, we treat high-suspicion

windows (2+, 3+, or 4+ flags) as segments where multiple proxies converge in the way commercial detectors are designed to target.

Ethical considerations

To align with academic standards and to protect the integrity of ongoing editorial processes, the full titles of the four manuscripts are not disclosed. Two manuscripts are already published, and two remain under consideration; withholding titles reduces re-identification risk, preserves the anonymity expected in peer review contexts, and prevents conflation of evaluative judgments with the present methodological investigation. This confidentiality does not affect reproducibility because all analyses operate on de-identified text segments, and the release package includes hashed document identifiers, section labels, and complete feature and outcome matrices.

Results

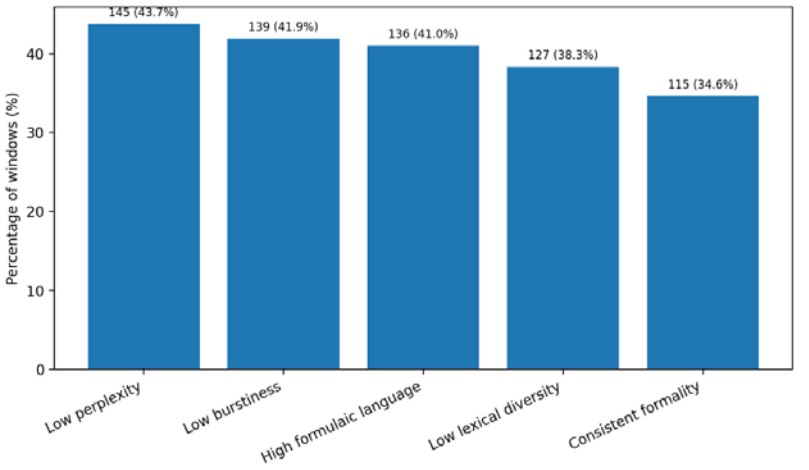
The four articles in the corpus contain 25,627 tokens and 3,350 distinct word types. Together, they comprise 404 paragraphs and 1,690 sentences, with an average sentence length of 15 words. None of the papers is an outlier in terms of length, sentence length, or paragraph count, indicating that they are structurally typical research articles rather than unusually short or long texts. After applying the 200-word sliding windows described in the methodology, the corpus yielded 332 analyzable windows: Technical Paper 1 (72 windows), Technical Paper 2 (107 windows), Linguistics Paper 1 (102 windows), and Linguistics Paper 2 (51 windows). Across this corpus, the AI detection simulator flagged 202 windows (60.8%) as suspicious at the baseline threshold (two or more flags), meaning they displayed at least some convergence of the five AI-like features. This broad pattern roughly mirrors the feedback from editorial offices and commercial tools.

Competing hypotheses of AI detection

Prevalence of individual features (prevalence model)

We first examined how often each of the five AI-detection proxies appeared across the 332 windows, treating each feature in isolation. The features occur at comparable rates, with no single proxy dominating the others. Low perplexity (predictable word sequences) appeared in 145 windows (43.7%), low burstiness (uniform sentence patterns) in 139 windows (41.9%), high formulaic language (recurrent multiword expressions) appeared in 136 windows (41.0%), low lexical diversity (restricted vocabulary range) appeared in 127 windows (38.3%), and consistent formality (little register variation) appeared in 115 windows (34.6%). Figure 1 illustrates the distribution of the identified proxies in the corpus.

Figure 1. Distribution of AI proxies in the sample



Source: Author's own elaboration.

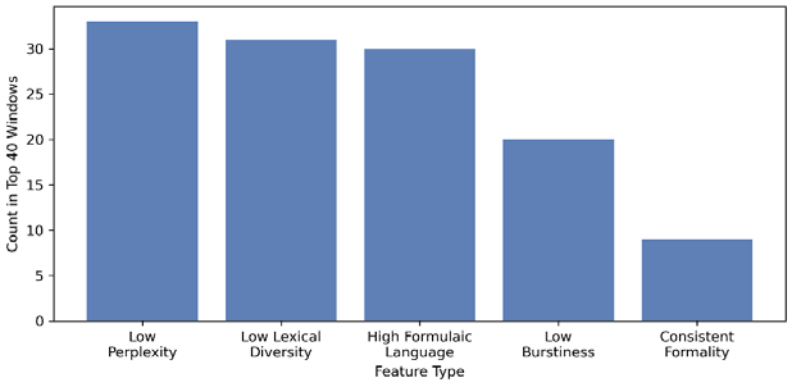
These results suggest that the corpus does not strongly resemble AI output when evaluated at the level of individual windows. As a prevalence model, this suggests that if detectors flagged the texts as AI-generated, they did so even though high concentrations of suspicious features are rare. Instead, the more plausible mechanism is that detectors treat any single feature crossing

a threshold as sufficient to raise suspicion, even if no other AI-like properties are present. If prevalence-based detectors treat any one of these features as a warning signal, then texts become vulnerable to flagging even when they exhibit only mild degrees of AI-like patterning.

Composition of high-scoring windows (clustering model)

We next turned to the clustering model, which focused on windows where multiple features co-occurred. Here, we examined the 40 most suspicious windows, defined as those with the highest cumulative scores, to see which combinations of features drive suspicion, and a combination of three proxies stands out. Low perplexity is present in 33 of 40 windows (82.5%), low lexical diversity appears in 31 windows (77.5%), and high formulaic language appears in 30 windows (75.0%). In contrast, low burstiness appears in 20 windows (50.0%), and consistent formality appears in 16 windows (40.0%). Figure 2 illustrates this pattern.

Figure 2. Clustering of AI proxies in the 40 most suspicious windows



Source: Author's own elaboration.

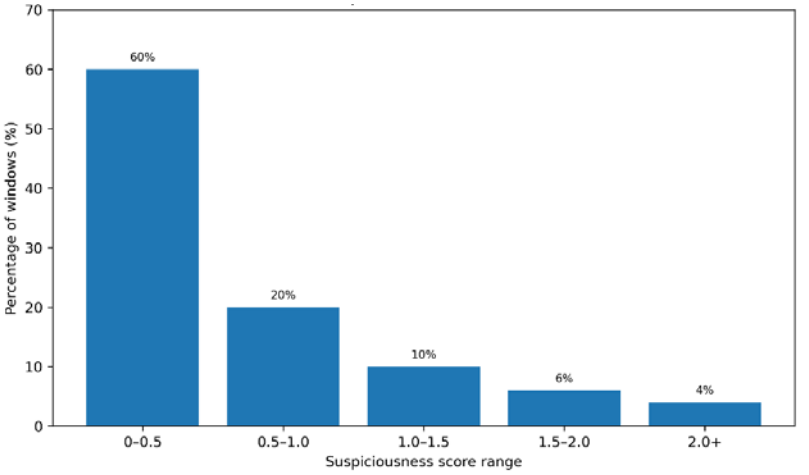
This pattern indicates that, when multiple features cluster together, three signals are doing most of the work. Predictable phrasing, heavy reuse of formulaic expressions, and restricted vocabulary are precisely the properties that prior research identifies both as core L2 academic strategies (e.g.,

formulaic bundles, conservative lexis, stable reporting frames) and as proxies used by AI detectors (low perplexity, high formulaicity, low lexical diversity). Crucially, the high-scoring windows do not occur at random in this data set. They are concentrated in sections that report statistical results, passages that describe standardized methodological procedures, and spans that present quantitative findings or parameter lists. In other words, the windows that look most AI-like to our simulator are also the most conventionalized parts of academic writing, where disciplines expect authors to use highly standardized wording. The simulator, therefore, flags segments that perform precisely the communicative functions that the IMRD structure requires, rather than passages that are stylistically unusual or semantically thin.

In our analysis, each article was divided into overlapping 200-word windows, and each window received a suspiciousness score from 0 to 5, reflecting how many of the five AI-detection proxies crossed their thresholds in that span. Scores of 0 or 1 are common and largely unremarkable, since a single flagged feature can arise from a local stylistic choice. Higher scores indicate the co--occurrence of multiple AI-like patterns in the same window, which is closer to the profile targeted by detection tools. Windows with two or more flags were treated as suspicious, with 2+ flags marking moderate suspicion, 3+ flags high suspicion, and 4+ flags very high suspicion.

However, the resulting score distribution is strongly right-skewed rather than bimodal. Figure 3 illustrates that most windows fall at the low end of the scale, and only a small subset accumulates three or more flags. This pattern indicates that conventional, highly formulaic passages form a long tail of high--scoring windows rather than a separate "AI" cluster, complicating any attempt to draw a hard categorical boundary between human and machine-like text.

Figure 3. Distribution of suspicious scores across windows



Source: Author's own elaboration.

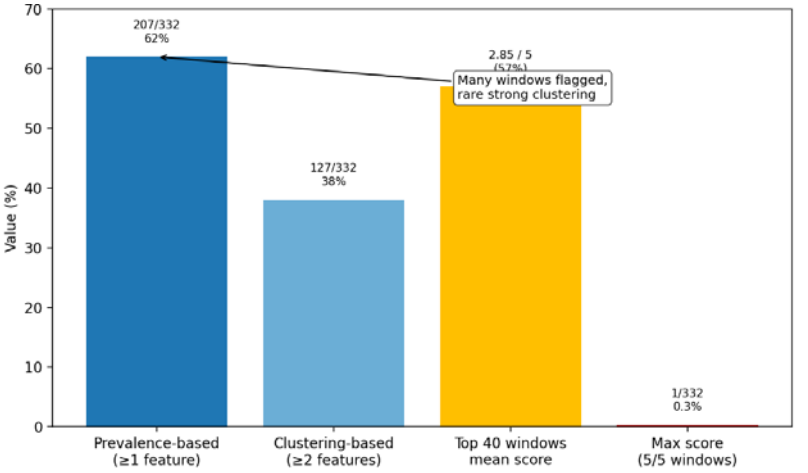
The relatively modest presence of consistent formality among high-scoring windows also suggests that register stability alone does not drive suspicion. Rather, it is the combination of predictable language, formulaic structure, and clustered terminology that produces the statistical profile most likely to be targeted by detection tools. Importantly, these combinations arise because the author follows disciplinary reporting norms, not because the text was generated by a language model.

The detection paradox

Under a prevalence-based model, which flags any window where at least one proxy exceeds its threshold, 62% of windows (207/332) would be treated as at risk. When we require at least two features in the same 200-word span, the clustering-based model still flags 38% of windows (127/332). Even within the 40 highest-scoring windows, the mean suspicious score is only 2.85 out of 5, and just one window in the entire corpus reaches the maximum of five co--occurring features. As visualized in Figure 4, this creates a detection paradox. Dense clustering of AI-like features is rare and mostly confined to genre--required segments, yet both architectures, especially the prevalence-based

model, would mark a substantial proportion of the corpus as suspicious. These findings sit against a striking practical backdrop, i.e., the texts in this corpus had already been rejected by multiple journals because they were "heavily AI--supported". Yet our analyses show that dense clustering of AI-like features is relatively rare, and when it does occur, it is typically confined to genre-required segments such as procedural descriptions, boilerplate method language, or parameter lists.

Figure 4. The detection paradox-flagging rates and score extremes



Source: Author's own elaboration.

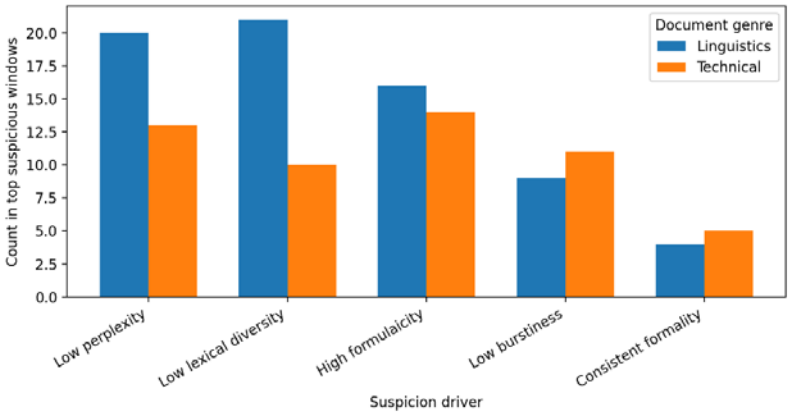
Commercial detectors and our simulator would flag a large share of these manuscripts as suspicious, particularly under prevalence-based logic, not because they contain long stretches of overtly machine-like prose, but because they are suffused with dispersed, individually common features that cross low thresholds. For multilingual scholars writing in English, this creates a structural vulnerability. L2 academic writers are precisely the group most likely to lean on conventional phrasing, fixed formulae, and other high-frequency academic routines to manage linguistic risk and conform to Anglophone norms. In doing so, they reproduce the very statistical signatures that current detectors interpret as AI authorship.

The analysis shows that the five features used in the AI detection simulator do not behave randomly in the corpus but follow a clear hierarchy in how often they appear and how strongly they contribute to suspicious scores. formulaic reuse, low perplexity, and low burstiness emerge as the main drivers of flagging. They appear in roughly 40–43% of all windows and concentrate in parts of the texts where academic conventions most tightly constrain wording, i.e., procedural descriptions, parameter listings, definitional and citation scaffolds, and standardized result statements. Low lexical diversity plays a secondary role, while consistent formality contributes more modestly.

Results by document genre

When we break the data down by document type, technical and linguistic papers show the same ordering of main suspicion drivers. Formulaic reuse, low perplexity, and low burstiness, followed by low lexical diversity and, finally, consistent formality. However, the relative emphasis of these drivers differs by genre. In the linguistics papers, suspicious windows show a slightly stronger contribution from formulaic reuse and low burstiness. This is consistent with sections that describe samples, annotation schemes, or coding procedures using templated phrases, and provide definitional and expository passages that recycle standard explanatory frames. In the technical papers, the same three drivers are evident, but consistent formality contributes less because the register is highly constrained from the outset. Suspicious windows here are dominated by standardized reporting frames (e.g., parameter estimations, model specifications), and parameter-dense prose where domain-specific terms repeat within narrow spans. Both patterns are hallmarks of legitimate discipline-specific writing, not symptoms of automated generation. Figure 5 visualizes the relative contribution of each proxy across the top 40 most suspicious windows by genre.

Figure 5. The most suspicious windows by genre



Source: Author's own elaboration.

Results by section

Finally, we examined where suspicious windows cluster within the IMRD structure. Although academic writing is conventional across all sections, suspicious flagging is not evenly distributed. In this corpus, Methods sections, and to a lesser extent Results, most frequently cross the suspicious thresholds. These sections are where low perplexity and low burstiness co-occur with high formulaic reuse, especially in procedural descriptions, stepwise enumerations, and parameter or variable listings. By contrast, introductions show moderate formulaic reuse and local dips in burstiness around definitional scaffolds and citation-heavy background passages, but these features rarely combine with enough intensity to trigger high suspicion scores.

Additionally, results in technical papers display elevated formulaicity and predictability in line with standardized statistical reporting, while linguistics results are more varied but still rely on recurring reporting frames. Discussions contain relatively few suspicious windows; when they do appear, they are more strongly associated with low perplexity and low burstiness than with heavy bundle reuse, reflecting tight but less templated interpretive prose. Abstracts and other short segments sometimes show low perplexity and low burstiness

owing to dense information packaging, but typically lack the sustained formulaic coverage seen in Methods. These sectional patterns align with the genre-typical functions of each part of the article. Procedural, definitional, and standardized reporting segments are predictably conventionalized and therefore more likely to resemble AI proxies, while interpretive segments are more varied and thus less likely to cross detector thresholds.

Discussion

Alignment with L2 academic writing and AI detection research

The distribution of features in this corpus is highly consistent with existing work on academic English, L2 writing, and World Englishes (Alfehaid & Alkhatib, 2023; Du & Hashimoto, 2023; Ellis et al., 2008; Li, 2024; Öztürk & Taşçı, 2023) Corpus and genre studies have long documented the heavy use of lexical bundles and formulaic sequences in Methods, Results, and definitional parts of Introductions, as well as in citation-dense stretches of text (Bajwa et al., 2020; Ellis et al., 2008; Öztürk & Taşçı, 2023). Genre analyses also predict highly regular staging in Introductions and standardized procedures in Methods and Results (Bajwa et al., 2020; Haq et al., 2023), which corresponds to lower burstiness and lower perplexity in those parts of the texts (Odri & Yoon, 2023; Picazo-Sanchez & Ortiz-Martin, 2024).

The findings also reinforce research on L2 academic writing that describes formulaic sequences as legitimate scaffolds for cohesion, clarity, and fluency, especially for multilingual writers (Öztürk & Taşçı, 2023). L2 authors are known to rely heavily on recurrent expressions, reporting frames, and definitional templates as they manage the dual demands of disciplinary content and an additional language. This reliance is neither an error nor a shortcut; it reflects successful socialization into the phraseology and routines of academic English (Showemimo, 2025; Giray, 2024). The concentration of suspicious scores in these routines supports the argument advanced in the literature review that L2 compensatory strategies and AI-detection proxies substantially overlap (Liang et al., 2023).

At the same time, the results align with empirical studies of AI detection tools, which show that current systems focus on the distributional properties of text (such as perplexity and burstiness) rather than on provenance, intent, or content knowledge (Aldeen et al., 2025). Prior evaluations have also noted that detectors' performance can vary by genre and section, with highly structured academic prose. For example, Popkov & Barrett (2025) found that ZeroGPT mis-labelled ≈62% of authentic behavioral-health articles as AI-generated, with the greatest errors in method-heavy sections. Similarly, Turnitin's AI-detector reported ≈4% false-positive rates for human-written sentences, but whole-document false positives rose sharply for formally structured essays (Showemimo, 2025). In L2 corpora, the problem intensifies. Studies of TOEFL essays (non-native English) show false-positive rates above 60% across seven popular GPT-detectors, because the limited lexical variability of learners' writing mimics the low-entropy signatures of machine-generated text (Liang et al., 2023).

Theoretical and practical implications

The findings highlight the detection paradox. Our simulator reveals that while individual AI-like features are common, strong clustering of multiple features is rare. Nonetheless, many windows can still be flagged under prevalence--based and clustering-based criteria due to low proxy thresholds used by detectors. This means that "more predictable" is mistakenly equated with "more likely AI-generated", contradicting insights from genre studies and corpus linguistics about academic prose. This dynamic creates a clear double bind for L2 writers. The sections where they are under the greatest pressure to follow disciplinary templates (Methods and Results) are also where our simulator finds the highest suspicious scores. Adhering closely to accepted reporting conventions increases the risk of being misclassified as using AI, but deviating from those conventions can undermine perceived credibility, clarity, and publishability. In this sense, the features driving high suspicious scores (low perplexity, low lexical diversity, and high formulaicity) are better understood as markers of successful participation in academic discourse communities than as signs of machine authorship.

This has several practical implications. For teaching and supervision, the findings underline the need for careful messaging around AI and academic writing. English for Academic Purposes (EAP) courses and supervisors should continue to promote appropriate conventional expressions and reporting frames, especially for multilingual writers, while also raising awareness of how detectors work and where they are likely to misfire. Non-native writers should not be discouraged from using formulaic language that signals disciplinary competence simply because it may resemble AI proxies.

For editorial policy and institutional practice, the results support treating AI detector outputs as screening tools rather than standalone evidence of misconduct. Editorial offices, universities, and funding bodies should be cautious about using raw detector scores in high-stakes decisions, particularly for L2 authors or submissions from diverse linguistic backgrounds. Where detectors raise concerns, follow-up should involve contextual judgement that considers genre, section, and writer background rather than defaulting to automatic sanction.

For AI detection research and development, the study adds empirical support to calls for more genre- and language-sensitive models. If false positives in this L2 corpus largely arise from normal academic conventions, then future detectors need to distinguish more carefully between predictable, formulaic text that is expected for a given genre and genuinely suspicious patterns. This could involve developing section-specific baselines that recognize the inherently higher conventionality of Methods and Results compared to Discussions, and incorporating information about text type, disciplinary genre, and possibly writer background so that expectations about predictability and formulaicity are calibrated rather than uniform, creating more equitable detection.

Conclusion

This proof-of-concept study combines a small, motivated sample of L2-authored research manuscripts with a transparent, detector-inspired simulator to explore one possible mechanism behind AI detection false positives. Within

this constrained setting, the most suspicious simulated segments tended to cluster in highly conventionalized methods and results prose characterized by low perplexity, low burstiness, and high formulaic reuse; precisely the kinds of patterns that are also encouraged by disciplinary writing norms and by legitimate L2 strategies for managing linguistic and cognitive load.

The study makes three main contributions. Theoretically, it refines the detection paradox by demonstrating how detectors can conflate linguistic predictability with machine authorship. Methodologically, it introduces a simulation-based, window-level approach that maps widely discussed detection heuristics onto authentic L2 research articles without relying on opaque commercial tools. Practically, the findings underscore the need for pedagogical guidance, editorial policies, and detector designs that recognize highly conventional academic language, especially in L2 writing, as a normal feature of scholarly communication rather than an automatic marker of AI misuse. The results suggest that protecting the integrity of scholarship will require not only better-calibrated detection architectures, but also institutional practices that allow multilingual authors to contextualize detector outputs and contest automated flags.

This work has limitations. The single-author design enables controlled analysis of detection mechanisms but constrains generalizability across L2 populations. The simulation approach, while transparent, does not directly test commercial detectors or compare false positive rates across tools. Future research should scale this framework to multi-author, multi-institution corpora spanning diverse L1 backgrounds, conduct prospective evaluations of section--weighted detection algorithms, and assess the impact of editorial interventions on false positive rates and perceived fairness in peer review.

References

Abd-Elaal, E.-S., Gamage, S. J. E., & Mills, J. E. (2022). Assisting academics to identify computer-generated writing. *European Journal of Engineering Education*, 47(5), 725–745. DOI: 10.1080/03043797.2022.2046709.

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, *31*(2), 81–92. DOI: 10.1016/j.esp.2011.08.004.

Ajideh, P., Zohrabi, M., & Ohbatalab, R. (2025). Stance markers in academic writing: Native vs. non-native (Iranian) authorship in hard and soft sciences research articles. *Critical Literary Studies*, *7*(2). DOI: 10.22034/cls.2025.63772.

Ahmed, J., & Ahmed, N. (2026). When Human Writing is Marked AI written: Investigating Turnitin's Misclassification of Students' Writing. *Research Journal for Social Affairs*, 4(1), 209–215. DOI: 10.71317/RJSA.004.01.0702.

Aldeen, S. D., Abbas, T., & Abbas, A. R. (2025). Review of Detecting Text generated by ChatGPT Using Machine and Deep-Learning Models: A Tools and Methods Analysis. *Diyala Journal of Engineering Sciences*, *18*(1), 34–54. DOI: 10.24237/djes.2025.18102.

Alfehaid, A., & Alkhatib, N. (2024). (Non)-Conformity to Native English Norms in Postgraduate Students' Writing in UK Universities: Perspectives of Native and Non-Native Students and Academic Staff. *International Journal of Arabic--English Studies (IJAES)*, *24*(1), 1–20. DOI: 10.33806/ijaes.v24i1.562.

Al Fattah, N. (2018). Cognitive load theory in the context of second language academic writing. *Higher Education Pedagogies*, *3*(1), 385–402. DOI: 10.1080/23752696.2018.1513812.

Albelihi, H. H. M., & Al-Ahdal, A. (2024). Overcoming error fossilization in academic writing: Strategies for Saudi EFL learners to move beyond the plateau. *Asian-Pacific Journal of Second and Foreign Language Education*, 9(75). DOI: 10.1186/s40862-024-00303-y.

Altakhaineh, A. R. M., Younes, A. S., & Allawama, A. (2024). A corpus--driven study of gratitude in English acknowledgements by Arabic-speaking

MA students: constructing L2 academic writer identity. *Cogent Arts & Humanities*, *11*(1). DOI: 10.1080/23311983.2024.2346361.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. DOI: 10.1145/3442188.3445922.

Bajwa, N. H., König, C. J., & Kunze, T. (2020). Evidence-based understanding of introductions of research articles. *Scientometrics*, *124*, 195–217. DOI: 10.1007/s11192-020-03475-9.

Caines, A., Benedetto, L., Taslimipoor, S., Davis, C., Gao, Y., Andersen, O. et al. (2023). On the application of large language models for language teaching and assessment technology. *arXiv:2307.08393*. DOI: 10.48550/arXiv.2307.08393.

Canli, Z., & Yağız, O. (2024). A contrastive investigation into the non-native speakers of English academicians' academic writing cognitions and challenges in the first and second languages. *Arab World English Journal*, *15*(1), 117–131. DOI: 10.24093/awej/vol15no1.8.

Casal, J. E., & Yoon, J. (2023). Frame-based formulaic features in L2 writing pedagogy: Variants, functions, and student writer perceptions in academic writing. *English for Specific Purposes*, *71*, 102–114. DOI: 10.1016/j.esp.2023.03.004.

Connor, U. (1996). *Contrastive rhetoric*. Cambridge University Press.

Doughman, J., Afzal, O. M., Toyin, H. O., Shehata, S., Nakov, P., & Talat, Z. (2025, January). Exploring the limitations of detecting machine-generated text. In: *Proceedings of the 31st International Conference on Computational Linguistics* (pp. 4274–4281). DOI: 10.48550/arXiv.2406.11073.

Du, Z., & Hashimoto, K. (2023). TCNAEC: Advancing sentence-level revision evaluation through diverse non-native academic English insights. *IEEE Access*, *11*, 144939–144952. DOI: 10.1109/ACCESS.2023.3342862.

Eid, F. M. S., & Mutahar, M. S. A. (2025). Bridging rhetorical differences: Arabic textual metaphors in academic writing and translation. *European Journal of Arts, Humanities and Social Sciences*, *2*(2), 183–201. DOI: 10.59324/ejahss.2025.2(2).21.

Ellis, N. C., Simpson-Vlach, R. I. T. A., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *42*(3), 375–396. DOI: 10.1002/j.1545-7249.2008.tb00137.x.

Flitcroft, M. A., Sheriff, S. N., Wolfrath, N., Maddula, R., McConnell, L., Xing, Y., & Kothari, A. N. (2024). Performance of artificial intelligence content detectors using human and artificial intelligence-generated scientific writing. *Annals of Surgical Oncology*, *31*(10), 6387–6393. DOI: 10.1245/s10434-024-15549-6.

Flowerdew, J. (2001). Attitudes of journal editors to nonnative speaker contributions. *TESOL Quarterly*, *35*(1), 121–150. DOI: 10.2307/3587862.

Ganjavi, A., & Conner, M. B. A. (2024). Publishers' and journals' instructions to authors on use of generative artificial intelligence in academic and scientific publishing: A bibliometric analysis. *BMJ*, 386, e077192. DOI: 10.1136/bmj-2023-077192.

Canagarajah, A. S. (2002). *A geopolitics of academic writing* (Vol. 163). University of Pittsburgh Press.

Gao, C. A., & Howard, F. N. S. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *NPJ Digital Medicine*, *6*(1), 75. DOI: 10.1038/s41746-023-00819-6.

Gehrmann, S., Strobelt, H., & Rush, A. M. (2019). GLTR: Statistical detection and visualization of generated text. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 111–116). Association for Computational Linguistics. DOI: 10.18653/v1/P19--3019.

Giray, L. (2024). The problem with false positives: AI detection unfairly accuses scholars of AI plagiarism. *The Serials Librarian*, *85*(5–6), 181–189. DOI: 10.1080/0361526X.2024.2433256.

Giray, L., Santos, K., & Manalang, F. (2025). Beyond policing: AI writing detection tools, trust, academic integrity, and their implications for college. *Internet Reference Services Quarterly*, *29*(1), 83–116. DOI: 10.1080/10875301.2024.2437174.

Gosselin, R. D. (2025). AI detectors are poor Western blot classifiers: A study of accuracy and predictive values. *PeerJ*, *13*, e18988. DOI: 10.7717/peerj.18988.

Granić, A., & Marangunić, N. (2019). Technology acceptance model in educational context: A systematic literature review. *British Journal of Educational Technology*, *50*(5), 2572–2593.DOI: 10.1111/bjet.12864.

Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J., & Wu, Y. (2023). How close is ChatGPT to human experts? Comparison corpus, evaluation, and detection. *arXiv:*2301.0759. DOI: 10.48550/arXiv.2301.07597.

Guo, H., Cheng, S., Zhang, K., Shen, G., & Zhang, X. (2025). CodeMirage: A multi-lingual benchmark for detecting AI-generated and paraphrased source code from production-level LLMs. *arXiv:2506.11059*. DOI: 10.48550/arXiv.2506.11059.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). Routledge. DOI: 10.4324/9780203431269.

Haq, Z. U., Naeem, H., Naeem, A., Iqbal, F., & Zaeem, D. (2023). Comparing human and artificial intelligence in writing for health journals: An exploratory study. *medRxiv*. DOI: 10.1101/2023.02.22.23286322.

Hinkel, E. (2001). Matters of cohesion in L2 academic texts. *Applied Language Learning*, *12*(2), 111–132.

Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features*. Routledge.

Hinkel, E. (2003). Simplicity without elegance: Features of sentences in L1 and L2 academic texts. *TESOL Quarterly*, *37*(2), 275–301. DOI: 10.2307/3588505.

Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing*, *31*, 58–69. DOI: 10.1016/j.jslw.2016.01.005.

Ippolito, D., Duckworth, D., Callison-Burch, C., & Eck, D. (2020). Automatic detection of generated text is easiest when humans are fooled. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1808–1822). Association for Computational Linguistics. DOI: 10.18653/v1/2020.acl-main.164.

Jarvis, S., & McNamara, D. S. (2012). Detecting the first language of second language writers using automated indices of cohesion, lexical sophistication, syntactic complexity, and conceptual knowledge. In: S. Jarvis & S. A. Crossley (Eds.), *Approaching language transfer through text classification: Explorations in the detection-based approach* (pp. 106–126). Channel View Publications. DOI: 10.21832/9781847696991-005.

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. Routledge.

Junina, A. K., Strauss, P. L., Wood, J. K., & Grant, L. (2025). A mixed-method inquiry into Arabic-speaking students' experiences with English academic writing at the undergraduate level. *Asia Pacific Journal of Education*, 45(1), 260–277. DOI: 10.1080/02188791.2022.2101986.

Kar, S. K., Bansal, T., Modi, S., & Singh, A. (2025). How Sensitive Are the Free AI-detector Tools in Detecting AI-generated Texts? A Comparison of Popular AI--detector Tools. *Indian Journal of Psychological Medicine*, 47(3), 275–278. DOI: 10.1177/02537176241247934.

Kashiha, H., & Chan, S. H. (2015). A little bit about: Differences in native and non-native speakers' use of formulaic language. *Australian Journal of Linguistics*, 35(4), 297–310. DOI: 10.1080/07268602.2015.1067132.

Kehkashan, T., Riaz, R. A., Al-Shamayleh, A. S., Akhunzada, A., Ali, N., Hamza, M., & Akbar, F. (2025). AI-generated text detection: A comprehensive review of methods, datasets, and applications. *Computer Science Review*, *58*, 100793. DOI: 10.1016/j.cosrev.2025.100793.

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge University Press. DOI: 10.1017/CBO9780511620607.

Laufer, B., & Waldman, T. (2011). Verb–noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*(2), 647–672. DOI: 10.1111/j.1467-9922.2010.00621.x.

Liang, W., Yükselgonul, M., & Mao, Y. J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4(7), 100779. DOI: 10.1016/j.patter.2023.100779.

Lillis, T. M., & Curry, M. J. (2010). *Academic writing in global context*. Routledge.

Liu, J. Q., & Hui, K. A. Y. (2024). The great detectives: Humans versus AI detectors in catching large language model-generated medical writing. *International Journal for Educational Integrity*, *20*(1), 8. DOI: 10.1007/s40979-024-00155-6.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36–62. DOI: 10.5054/tq.2011.240859.

Mavrou, I., & Chao, J. (2023). What does linguistic distance predict when it comes to L2 writing of adult immigrant learners of Spanish? *Written Communication*, *40*(3), 943–975. DOI: 10.1177/07410883231169511.

Miralles-González, P., Huertas-Tato, J., Martín, A., & Camacho, D. (2025). Not all tokens are created equal: Perplexity Attention Weighted Networks for AI generated text detection. *arXiv:2501.03940*. DOI: https://doi.org/10.48550/arXiv.2501.03940.

Mitchell, E., Lee, Y., Khazatsky, A., Manning, C. D., & Finn, C. (2023). *DetectGPT: Zero-Shot Machine-Generated Text Detection using Probability Curvature* (Version 2). *arXiv:2301.11305*. DOI: 10.48550/ARXIV.2301.11305.

Mu, C., & Zhang, L. J. (2018). Understanding Chinese multilingual scholars' experiences of publishing research in English. *Journal of Scholarly Publishing*, *49*(4), 397–418. DOI: 10.3138/jsp.49.4.02.

Muñoz-Ortiz, A., Gómez-Rodríguez, C., & Vilares, D. (2024). Contrasting linguistic patterns in human and LLM-generated news text. *Artificial Intelligence Review*, *57*(10), 265. DOI: 10.1007/s10462-024-10903-2.

Murdoch, Y. D., Lim, H., & Cho, J. (2021). Writing center visitors: Influence of L1 writing skills on students' exophonic writings. *SAGE Open*, *11*(4). DOI: 10.1177/21582440211062234.

Odri, G. A., & Yoon, D. J. Y. (2023). Detecting generative artificial intelligence in scientific articles: Evasion techniques and implications for scientific integrity. *Orthopaedics & Traumatology: Surgery & Research*, *109*(8), 103706. DOI: 10.1016/j.otsr.2023.103706.

Opara, C. (2025). Distinguishing AI-generated and human-written text through psycholinguistic analysis. In: *International Conference on Artificial Intelligence in Education* (pp. 212–219). Cham: Springer Nature Switzerland. DOI: 10.48550/arXiv.2505.01800.

Öztürk, Y., & Taşçı, S. (2023). A corpus-based analysis of lexical bundles in non-native post graduate academic writing and a potential L1 influence. *REFLections*, *30*(2), 488–505. DOI: 10.61508/refl.v30i2.267463.

Pan, F. (2018). A multidimensional analysis of L1–L2 differences across three advanced levels. *Southern African Linguistics and Applied Language Studies*, *36*(2), 117–131. DOI: 10.2989/16073614.2018.1476162.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, *32*, 130–149. DOI: 10.1017/S0267190512000098.

Pedersen, A. M. (2011). Writing Across Languages, Disciplines, and Sources: Second Language Writers in Jordan. *Across the Disciplines*, *8*(1), 109–119. DOI: 10.37514/ATD-J.2011.8.1.06.

Picazo-Sanchez, P., & Ortiz-Martin, L. (2024). Analysing the impact of ChatGPT in research: P. Picazo-Sanchez and L. Ortiz-Martin. *Applied Intelligence*, *54*(5), 4172–4188. DOI: 10.1007/s10489-024-05298-0.

Popkov, A. A., & Barrett, T. S. (2025). AI vs academia: Experimental study on AI text detectors' accuracy in behavioral health academic writing. *Accountability in Research*, *32*(7), 1072–1088. DOI: 10.1080/08989621.2024.2331757.

Shin, Y. K. (2019). Do native writers always have a head start over nonnative writers? The use of lexical bundles in college students' essays. *Journal of English for Academic Purposes*, *40*, 1–14. DOI: 10.1016/j.jeap.2019.04.004.

Showemimo, P. (2025). The AI Paradox in Academia: When Writing Too Well Raises Red Flags. *SSRN*. DOI: 10.2139/ssrn.5255976.

Shchemeleva, I. (2021). There's no discrimination: The association between journals, title rephrasing of the given reviewer suggestions, prevalence of text-recycling, and publishers' policies in English. *Publications*, *9*(1), 8. DOI: 10.3390/publications9010008.

Shehata, A. M. K., & Eldakar, M. A. M. (2018). Publishing research in the international context: An analysis of Egyptian social sciences scholars' academic writing behaviour. *The Electronic Library*, *36*(5), 910–924. DOI: 10.1108/EL-01-2017-0005.

Subandowo, D., Sárdi, C., & Thresia, F. (2025). An investigation of English academic writing strategies employed by Indonesian graduate students in an English medium instruction (EMI) context. *Asian-Pacific Journal of Second and Foreign Language Education*, *10*(1), 38. DOI: 10.1186/s40862-025-00345-w.

Taşçı, S., & Öztürk, Y. (2021). Post-predicate that-clauses controlled by verbs in native and non-native academic writing: A corpus-based study. *Australian Journal of Applied Linguistics*, *4*(1), 18–33. DOI: 10.29140/ajal.v4n1.486.

Tufts, B., Zhao, X., & Li, L. (2025). A practical examination of AI-generated text detectors for large language models. In: *Findings of the Association for Computational Linguistics: NAACL 2025* (pp. 4824–4841). DOI: 10.48550/arXiv.2412.05139.

Uchendu, A., Le, T., Shu, K., & Lee, D. (2020). Authorship attribution for neural text generation. In: *Proceedings of the 2020 Conference on Empirical Methods*

in Natural Language Processing (EMNLP) (pp. 8384–8395). Association for Computational Linguistics. DOI: 10.18653/v1/2020.emnlp-main.673.

Valipour, V., Assadi, N., & Asl, H. D. (2017). The generic structures and lexico-grammaticality in English academic research papers. *Southern African Linguistics and Applied Language Studies*, *35*(2), 169–182. DOI: 10.2989/16073614.2017.1373365.

Wang, X., Zhang, W., & Rajtmajer, S. (2024). Monolingual and multilingual misinformation detection for low-resource languages: A comprehensive survey. *arXiv:2410.18390*. DOI: 10.48550/arXiv.2410.18390.

Weber-Wulff, D., Anohina-Naumeca, A., & Bjelobaba, S. L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, *19*(1), 1–39. DOI: 10.1007/s40979-023-00146-z.

Won, D., Shin, Y. K., Kim, H. & Yoo, I. W. (2025). Advancing Language Assessment with GPT: Is It Nonnative-Language Friendly? *Language Assessment Quarterly*, *1*, 1–18. DOI: 10.1080/15434303.2024.2444349.

Zaini, A., & Ollerhead, S. (2019). Reverse contrastive rhetoric in expository writing: Transfer and power relations at work. *Southern African Linguistics and Applied Language Studies*, *37*(1), 41–61. DOI: 10.2989/16073614.2019.1609364.

Appendix A

This appendix provides a summary of the five proxies coded in this study, along with their computational operationalization and flagging criteria. All coding, feature extraction, and analysis were conducted programmatically in Python. Complete technical specifications, including software libraries, mathematical formulas, and reproducibility documentation, are available in the full Technical Codebook upon request.

Coding scheme

Feature	Definition	Computational Measure	Flagging Threshold
F1: Low Perplexity	Text predictability to language model	GPT-2 perplexity score	≤ 25th percentile
F2: Low Burstiness	Uniformity of word probability	Variance of GPT--2 log probabilities	≤ 25th percentile
F3: High Formulaic Language	Density of recurrent multi--word sequences	Bundle coverage (3-5 grams, 3+ occurrences)	≥ 75th percentile
F4: Consistent Formality	Uniformity of register across sentences	Variance of Heylighen--Dewaele formality scores	≤ 25th percentile
F5: Low Lexical Diversity	Repetitiveness of vocabulary	HD-D lexical diversity score	≤ 25th percentile

Appendix B

We present four examples representing different flag patterns across genres and sections. Example B is extracted from a technical paper. Example B is an excerpt from a technical paper coded as "Technical 1". The text is extracted from the results section. This extract represents Window 46, which received a Suspicious Score of 5/5, Flagging features F1, F2, F3, F4, F5 (consult coding scheme). Here is the excerpt:

> "enables a deeper understanding of the model's predictive and explanatory power, which will be discussed in the following section. To assess the significance of the path coefficients, bootstrapping with 5,000 sub--samples was conducted in SMART PLS 4.0. The path analysis revealed complex relationships..."

The specifications of the flagged window are reported here for further illustration. This window achieves a perfect 5/5 score through extreme z-scores

across all features. The text is highly predictable (z = -1.88 perplexity), extremely smooth (z = -2.97 burstiness), formulaic ("path coefficients", "bootstrapping", "sub-samples"), maintains consistent formal register (z = -0.84), and uses limited vocabulary (z = -2.38 lexical diversity). This pattern is genre-appropriate for technical results reporting with statistical notation and methodological terminology.

FEATURE PROFILE:

F1 (Low Perplexity):	FLAGGED (z = -1.88, PPL = 364.2)
F2 (Low Burstiness):	FLAGGED (z = -2.97, Burst = 1.64)
F3 (High Formulaic):	FLAGGED (z = 2.06, Coverage = 0.176)
F4 (Formality Variance):	FLAGGED (z = -0.84, Var = 0.175)
F5 (Low Lexical Diversity):	FLAGGED (z = -2.38, HD-D = 0.783)

Example C is an excerpt from a technical paper coded as "Technical 1". The text is extracted from the methods section. This extract represents Window 16, which received a Suspicious Score of 2/5, Flagging features F2, F4 (consult coding scheme). Here is the excerpt:

> "understand how students engage with educational technologies (Granić & Maranguni, 2019). In the realm of language teaching and learning, TAM has been employed to examine the importance of technology acceptance among both educators and students. For educators, accepting technology is essential for..."

The specifications of the flagged window are reported here for further illustration. This window achieves a 2/5 score. This methodological description shows moderate flagging (2/5). The text maintains a consistent formal register (F4 flagged, z = -2.25) and smooth flow (F2 flagged), but perplexity is actually high (z = 1.26), indicating varied sentence structures and conceptual explanations. This is consistent with the possibility that not all technical writing triggers all

flags; the literature review and methods sections show more linguistic variation than pure results reporting.

FEATURE PROFILE:

F1 (Low Perplexity):	NOT FLAGGED (z = 1.26, PPL = 568.8)
F2 (Low Burstiness):	FLAGGED (z = -0.83, Burst = 2.30)
F3 (High Formulaic):	NOT FLAGGED (z = -1.30, Coverage = 0.081)
F4 (Formality Variance):	FLAGGED (z = -2.25, Var = 0.154)
F5 (Low Lexical Diversity):	NOT FLAGGED (z = 0.16, HD-D = 0.853)

Example D is an excerpt from a technical paper coded as "Linguistics 2". The text is extracted from the literature review section. This extract represents Window 24, which received a Suspicious Score of 3/5, Flagging features F3, F4, F5 (consult coding scheme). Here is the excerpt:

> "reflects the overall acceptability of the pragmatic functions linked to the religious marker … in everyday spoken Jordanian Arabic (SJA). This benchmark precedes the analysis of how face-related factors influence acceptability rates. To capture these preferences, a 5-point L…"

The specifications of the flagged window are reported here for further illustration. This window achieves a 3/5 score. This theoretical framework section shows a different flag pattern (F3, F4, F5 without F1/F2). The text is formulaic (repeated conceptual terminology: "acceptability", "pragmatic functions", "face-related factors") and maintains consistent formality, but perplexity and burstiness are NOT flagged because the author introduces new concepts and examples (phonetic transcription, language-specific terms). This illustrates how theoretical exposition differs from statistical reporting.

FEATURE PROFILE:

F1 (Low Perplexity):	NOT FLAGGED (z = -0.61, PPL = 437.8)
F2 (Low Burstiness):	NOT FLAGGED (z = 1.04, Burst = 2.77)
F3 (High Formulaic):	FLAGGED (z = 1.47, Coverage = 0.179)
F4 (Formality Variance):	FLAGGED (z = -1.01, Var = 0.003)
F5 (Low Lexical Diversity):	FLAGGED (z = -0.83, HD-D = 0.764)

Our conclusions are constrained by the small, personal nature of the corpus, the focus on a single L1 background and disciplinary area, and the use of a simulator rather than direct access to commercial detectors. All analyses are cross-sectional and observational and thus cannot establish causal relationships between specific textual properties and detector behaviour. Future research with larger, more diverse samples and direct access to detection outputs is needed to test and refine the hypotheses advanced here.

DOI: 10.2478/doc-2026-0004

This is an open access article licensed under the Creative Commons AttributionNonCommercial-NoDerivatives 4.0 International (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Irena Vassileva

Department of Foreign Languages and Cultures, New Bulgarian University, Sofia, Bulgaria

ivasileva@nbu.bg

ORCID ID: 0000-0002-0150-6375

Mariya Chankova

Department of Foreign Languages, South-West University Neofit Rilski, Blagoevgrad, Bulgaria

mariya.chankova@swu.bg

ORCID ID: 0000-0003-3017-125X

GenAI and Authorship in Academia: The (Im)Possible Creator. An Integrative Review

Article history:

Received **29 October 2025**

Revised **15 January 2026**

Accepted **19 February 2026**

Available online **23 June 2026**

Abstract: Amid vivid ethical discussions on using AI to assist scholars in the writing process, this contribution tackles the question about Generative AI and authorship from a philosophy of language and academic perspective. In this integrative review, we look into the ways Generative AI impacts the very notion of authorship on the stages of knowledge production, knowledge consumption and knowledge dissemination. We focus on the implications of using Large Language Models with their enhanced capabilities of generating convincing imitations of human natural language for our conception of authorship. Contextualizing this capability to the realm of academic production, we seek to explore whether generative AI is going to further add to the postmodern destruction of the author, and how it relates to a core aspect of academic integrity, namely acknowledging the intellectual contributions of other scholars.

Keywords: GenAI, authorship, academic integrity, knowledge production, knowledge consumption

Introduction

LLMs and generative AI tools are everywhere: submerging content creation online (Chankova & Vassileva, 2024), impacting deeply the very fabric of academic publishing (Perkins & Roe, 2024), interfering with learning on every imaginable level of autonomous internet users (Yu, 2023), shaping up pedagogical practices and overrunning didactic material preparation, formative assessment and feedback (Nguyen et al., 2024). In order to position the discussion of generative--AI tools in the framework of text-creating activities, we look into the question

of how our conception of authorship in academia is being impacted by these tools. Two primary applications are envisioned as a backdrop of the discussion: first, academic publishing and the guidelines to acknowledge the generative-AI tools used in the preparation of the manuscript (according to COPE, 2024); and second, the more general question of the use of generative-AI tools in academics' production. Both perspectives seek to answer the question of which uses of generative-AI tools can be described as ethical in terms of writing assistants.

The LLMs which made possible the elaboration of the ChatGPT-like algorithm have fascinated users with several particularities: they can imitate patterns of human speech; utilizing a larger window of local dependencies, they can produce speech that appears coherent across paragraphs (Bender et al., 2021); they can perform tasks of language manipulation, such as paraphrasing, general editing, rewriting (Lingard, 2023). The vast amount of data that is being used in the training (the process of feeding more linguistic patterns into the algorithm) allows these systems to produce on-topic text generations immediately, creating the illusion that they respond to the human user. Users have quickly discovered that these systems are prone to hallucinations and are quite plainly generic in their output (van Dis et al., 2023; Xames & Shefa, 2023; Grimaldi & Ehrler, 2023). These features stem directly from the algorithm, which is based on a randomized selection of linguistic patterns: since the algorithm lacks semantic understanding of the language data it manipulates, it tends to produce random item associations which make no coherent sense and/or express false propositions about the world (Bender et al., 2021; Titus, 2024). Even though hallucinations rates have been significantly reduced in later models (see OpenAI, 2025), this is deemed to be a structural feature of Large Language Models, and, in this sense, inevitable (Xu et al., 2024; Xu, 2025). This is one of the reasons why it can be argued that these algorithms cannot replace information search: they produce a probability generation that completely obscures the sources of the ideas one interprets them to contain. This very same particularity of the algorithm could make it a useful tool to jumpstart the creative process itself, taking on the role of a brainstorming session with outlines, ideas, summaries, which can then be exploited and developed by the user (Lingard, 2023; Grimaldi & Ehrler, 2023).

Two more brushstrokes to complete the context setting. Importantly, in the absence of compelling evidence of AI's ability to have communicative intention, it cannot be described as communicating and thus does not reflexively intend to create in the human user an understanding of the communicative intention: this particularity has led Bender and Hanna to call language models "synthetic text extruding machines" (2025, p. 31), underlining the danger of mistaking human's capacity for interpreting the textual output of GenAI for GenAI's thinking, understanding or feeling. This idea will be developed in a subsequent section of this article.

Secondly, the predictions of a fundamental, drastic change in the very fabric of the workforce with the arrival of generative AI systems (for example, McKinsey & Company, 2023) a couple of years into their adoption in various domains of human activities take a particular shape in reality. A large-scale Danish study by economists investigated 25 000 workers in 11 different sectors that were considered susceptible to be affected the most by the introduction of AI systems and the impact of AI use on their professional activities and found an average of 2.8% of time gains due to AI introduction (Humlum & Vestergaard, 2025). It is further emphasized that these gains are affected by an average increase in work charges as people now have to complete tasks that they did not have before, such as verification of AI output, this charge being between 8% for accounting to 10% for the educational sector, up to 15% increase in work tasks for law. Of course, the adoption rate of AI is different for different sectors, and many factors influence these decisions in the industries, not to mention the fact that different sectors will be impacted differently by the adoption of these tools. More broadly, preliminary findings of an MIT study entitled *The State of AI in Business 2025* uncovered an important divide in returns for investment for companies that invest in AI—according to the report, 95% of the companies with important investments in the development and implementation of AI technologies receive no return on investment (Challapally et al., 2025). These findings suggest that the implementation and exploitation for profit of these tools is still largely under way, even though on the individual level, they may be evaluated as having a measurable impact on productivity and/or efficiency.

Our preoccupation here concerns academia only, which leads us to another problem, namely the evaluation of the impact AI systems may have on the state of academic production. To quote Bender and Hanna, who spoke of workflow automation using ChatGPT-like systems: "We find ourselves asking: if they couldn't be bothered to write this, why should we be bothered to read it?" (Bender & Hanna, 2025, p. 52) This sentiment applies fully, in our view, to the potential AI bearing on academia. In this contribution, we set out to explore the question of how GenAI tools have impacted and can potentially impact our understanding of authorship.

Aims and methods

In what follows, we turn to the notion of authorship, which we will tackle from a philosophy of language and an academic perspective. We adopt the approach of the integrative review, as it allows us to draw ideas from multiple and diverse sources in order to integrate them into a holistic, coherent view on the problem of authorship in the age of GenAI. This approach permits a problematization of AI-augmented authorship from a multiple aspect perspective. We operate within the context of a GenAI-saturated academic environment, where scholars, instructors and students alike have largely adopted GenAI tools for various purposes (Chankova, in print), navigating the complex challenges these tools can potentially create for academia (Hughes et al., 2025). Our discussion of GenAI impact on academia will be limited to Large Language Models for their potential applications in writing research (language and style enhancement, structuring and planning, draft writing, idea generation, etc.). We do not consider other types of AI, which may have a vast array of applications in a multitude of domains. Our investigation is best framed in terms of ethical and responsible GenAI use, its high resource-intensiveness notwithstanding.

The analysis is organized as follows: we start by considering authorship and the construction of knowledge (section Authorship and knowledge creation), then we explore the ideas suggested by publishers in an effort to uphold academic

integrity in publishing research (section Authorship and knowledge distribution), and finally we turn to authorship and the consumption of knowledge (section Authorship and knowledge consumption), as we suggest that the primary aim of academic communication is the creation, dissemination and consumption of knowledge.

Authorship and knowledge creation

Historical perspectives

Historically, the concept of scientific language as a "universal code" stripped of individuality progressively gave rise to a discourse marked by strong formalization and standardization. This extended beyond grammar, vocabulary, and textual organization to shape how authors present themselves and how they engage with their readers. At the same time, the construction, production, and dissemination of knowledge are firmly embedded in culture-specific norms that have developed historically. As a result, distinct intellectual styles of presenting scientific discourse have emerged.

Rosenblatt (2018, p. 466) argues that human writing and reading are fundamental to the creation of meaning. These processes entail interactions not only with others—authors and readers—but also with oneself, drawing on memory and personal experience. Such dimensions cannot be replicated by AI. The process of writing is a function of the author's experience and individuality, participating in its turn in the construction of society, helping to articulate interpersonal relations, perceptions and representations (also Bazerman & Paradis, 1991, p. 3).

Two long-standing but still highly influential philosophical approaches to knowledge creation have to be taken into consideration here. First, according to social constructionists, knowledge is shaped by, and relative to, the community in which it is produced; meaning and understanding depend on a shared conceptual framework expressed through social norms. This makes it possible to define an 'academic discourse community' by identifying its underlying norms. Social constructionism is philosophically grounded in the work of analytic thinkers

such as Frege, Russell, and Chomsky, but it also resonates with speech-act theorists like Searle and phenomenologists such as Gadamer.

In contrast, externalists adopt an anti-Cartesian stance, drawing on philosophical traditions associated with Nietzsche, Dewey, Heidegger, Davidson, Derrida, and others. They view propositional attitudes as emerging from public interaction, where knowledge is achieved through interpretation rather than through the mediation of a prior framework of social norms. Communication, therefore, relies on deviation—or what Bakhtin terms *addressivity*, i.e. the recognition of the otherness of the other, if we desire to communicate at all. For Bakhtin, polyphony is inherent in all words or forms (1981, p. 293), i.e., it is impossible to create a totally new text.

GenAI tools are based on the social-constructionist understanding of knowledge creation, since LLMs draw on existing textual databases and extract repetitive, thus also normatively shaped textual strings. They therefore ignore the possibility of interpretation of information presented in a way deviating from the said norms. The latter, on their part (at least for the time being), are primarily centered around the Anglo-American views about knowledge representation and hardly allow for textualizations based on different rhetorical styles in academia.

Consequently, intercultural communication poses even greater challenges for the author, who is expected to navigate between different systems of cultural values and beliefs, sometimes entirely unfamiliar ones. This challenge arises not only when writing in a foreign language but also when addressing international audiences. It entails, as Bizzell (1990, p. 53) puts it, "going native", that is, reshaping one's identity as constructed by national culture, society, and the conventions of the native scientific community. The difficulty of fully transforming one's identity becomes particularly acute for scholars from cultural backgrounds that differ considerably from the target context, or for those who, due to political or economic isolation, have long been excluded from the international exchange of knowledge. Cultural identity and the relationship between the individual and society, however, can vary considerably. Foley summarises the Western concept of person as "individualism" where the "basic premise of this ideology is egocentric personal autonomy" (1997, p. 265). Within an egocentric, individualist ideology, society is positioned as secondary to the individual. Inequality in such

contexts is viewed as the outcome of free competition among individuals who are, at least in theory, considered equal. By contrast, in many traditional cultures, the concept of personhood is fundamentally different: meaning is generated through interaction and achieved through the active participation of all involved. Here, the audience is not a passive recipient expected merely to grasp the intentions of a fixed speaker (for a detailed discussion on the development of the philosophical views of authorship see Vassileva, 2006).

Reframing these aspects of the writing process and knowledge creation and transmission in the augmented author perspective, it appears that AI-produced output glosses over every single aspect we mentioned: such output does not in any meaningful way relate to forming propositional attitudes by an individual with their specific experience, perceptions and interpretations, neither does it reflect in any meaningful way the cultural baggage the human may carry with them, nor does it carry over authorial voices—or permit their development and polishing through the practice of the craft of writing.

Current views of human–GenAI interaction

From present-day perspective, it seems that the deconstruction of authorship, or the death of the author, to use Barthes' phrase, which centered around the destruction of the idea that a text possessed necessarily one ultimate interpretation (the message of "Author-God"), accommodates well the promise of another type of intelligence that seems capable of generating texts that simulate human production. It also reflects Barthes' conception of text as this "multidimensional space, in which a variety of writings, none of them original, blend and clash" (1977, p. 146). This description fits the AI algorithm really well: its training on vast amounts of texts, its capability of processing thousands of signs in the computational process of executing the selection algorithm in response to the prompt, all of those tally with the word mash which is then interpreted using the same sense-making defaults as the output of a human author. The output generated by ChatGPT is compared to a kind of "generalized view of people who write about a particular topic" (Van Woudenberg et al., 2024, p. 20), while others are at pains to describe such output as qualifying as

propositions in the first place (Frye, 2023, editor-in-chief's foreword). Indeed, without characterization, contextualization, source attribution, explicitation of connections, and the reference act (as a shortcut, we refer to Searle's notion of meaningfully realizing the act of referring to objects and realities in the world), it is difficult to evaluate the output in terms of its ontological value, even if it sometimes can be evaluated as true or false due to independent verification (also pointed out by Van Woudenberg et al., 2024).

The question of evaluating the AI systems' output appears to be settled by the very particularities of the LLM's system: it is by chance that output sentences may turn out to be true, not by design (Van Woudenberg et al., 2024, p. 5). To Bakhtin's definition of text, Treharne and Willan (2019) oppose another, one that takes into account the intentional and deliberate social move that motivates its creation to purposefully convey a message that has interpretable meaning. This echoes Bender and Hanna's insistence that 'normal' writing involves intentionality (in Brentano's sense), and writers, as do speakers, usually have illocutionary and perlocutionary intentions (Searle, 1969). The following definition of the notion of author has been suggested recently:

> "At root, an author is someone who engages in normal writing, and normal writing embeds normative, agency embellishing capacities, including certain kinds of belief, intentions, and responsibilities, like the ability to participate in practices in which one could be held accountable overtime for their words (ledger function); or would be answerable for specific claims made (answerability function); or participate in practices of blame and praise (desert function); or practices surrounding the seriousness with which one should treat certain statements (gravity discerning function); or practices of providing compensation, when necessary (compensation function)" (Van Woudenberg et al., 2024, p. 10).

Thus, authorship is a lasting relationship that the author entertains with their creation, actively serving as a reminder of the accountability and responsibility of each individual author for the ideas put forth, as well as serving the function of social (or academic) recognition. The text would serve

as sharing both the results and the evidence of the coming to know process, deliberately undergone by the author.

AI tools' users tend to interpret the output as if it had these attributes (on perceived AI agency, see Hwang & Won, 2022; on perceived agency and trust, see Vanneste & Puranam, 2024), which accounts for the perception of these tools as human substitutes: this is the default interpretation. Similar considerations have led legal scholars to rule on the legal status of machine-generated production by suggesting it should not be protected by copyright (Gervais, 2020). Moreover, encouraging machine-generated artistic production is seen as seriously threatening human progress (Gervais, 2020; Bender & Hanna, 2025).

Goffman's (1981) model of the author in terms of the production end of communication encompasses the following participant roles: the author, the one scripting the lines; the principal, the one endorsing the message; and the animator, the one vocalizing the message (the sound box). In academia, where ghost-writing is proscribed as practice, the researcher ideally embodies all three of the roles, as it is their responsibility that is engaged when they are involved in academic communication. Line-scripting is supposed to reflect authorial voice and style idiosyncrasies, ideas, stances and beliefs. It reflects the result of the thought process: thinking and writing are fundamentally inseparable processes (Bender & Hanna, 2025, p. 50). Using a text-generating system to generate the lines resembles in substance to ghost-writing, a practice in which the responsibility is claimed by the principal for content created by somebody else.

Ghost-writing (or contract cheating) is deemed an infringement of academic integrity—students (or academics, for that matter) who resort to buying essays/articles ghost-written for them by a third party are guilty of academic dishonesty. This may be used as validation for the presupposition still in play in academic writing and integrity matters that the processes of writing and thinking are interconnected and are a reflection of the special relationship between content and author. In any case, outsourcing a crucial part of the knowledge-creating process seems to be counterintuitive to the realization of human intentionality. Yet, GenAI tools' use is tolerated to a certain extent in the academic community.

The "distant writing" hypothesis in academic communication

The notion of "distant writing" was introduced by Floridi (2025) to describe a creative literary writing practice that is conjointly performed by a human agent acting as the designer of the narrative and by LLMs acting as writing generators. It will be discussed here in order to evaluate the possibilities of its application to the production and consumption of academic texts. The theoretical background of distant writing integrates ideas from multiple fields. From modal logic and narrative theory, it adopts the view of narrative as a boundless space of possible worlds constrained by coherence. From design theory, it makes use of the idea of creativity as an exploration process. From computational theory, it incorporates the human-computer partnership as a source of novel possibilities. Beyond these, it introduces the principle of "isotropy of the narrative space", meaning that narratives can develop in any direction as long as coherence is preserved, challenging traditional notions of fixed narrative progression.

Discussing the merits of such a theoretical construct for literary creation is beyond our focus in this integrative review; however, academic texts as a rule have pre-defined aims and outcomes, with highly standardized paths of progression, so that the space they move in is further constrained by disciplinary norms and expectations. Therefore, this space cannot possibly have an isotropic character, that is, it is not 'boundless'.

The methodology of distant writing treats composition as designing prompts that direct an LLM to produce the desired narrative. The mechanics of writing thus becomes a process of systematic prompting and refinement, where the quality of the output depends on the precision of the input. Through iterative progressive refinement, the author shapes the text and ultimately assumes responsibility for the final product. A crucial point here is the distinction made between *responsibility* and *ownership*, stemming from the 'decoupling' of the traditional functions of the author as both the conceptualizer and the physical writer of the text, assuming the role of operator, a kind of 'meta-author', directly involved in mapping out the plot, but not in the writing process itself (Floridi, 2025, p. 9).

The challenge of losing the individual author's voice, typically associated with unique stylistic figures, rhetorical patterns and language choices that

contribute to the recognition and attribution of literary texts to specific writers is not deemed serious from the perspective of distant writing: as Floridi (2025, p. 16) points out, the lack of mental states and thus intentional choices on the part of LLMs does not impede the development of unique "wrAIting styles" which go beyond the stylistic limitations that may be contained in the prompts. Transposing these considerations to the creation of academic texts, it could be suggested that although the notion of 'voice' is also one of their critical features, its realization is presupposed by completely different norms and expectations, while the style generally demonstrates recurrent structures and lexical choices typical of the respective discipline which, curiously, may make such texts more suitable for machine production.

Another concept that needs reassessment within Floridi's framework is that of *originality*, which has been impacted by a long-standing tradition of intertextuality theories describing texts as "mosaics of quotations"—for Floridi, distant writing underscores and systematizes intertextuality (2025, p. 19). Besides, while originality in fiction lies both on the plane of ideas and their linguistic expression, in academic texts it mainly concerns the novelty of ideas, supporting results and discoveries or inventions. Intertextuality, on its part, is expected to be clearly indicated through attribution, in contrast to its subtle character in literary texts. From an ethical viewpoint, Floridi sees the advent of distant writing as complicating "attribution and accountability", especially considering existing legal frameworks regulating intellectual property and copyright law "that may be inadequate for collaborative human-AI creativity" (2025, p. 26).

More broadly, this philosophical framework challenges the belief that meaningful texts must be purely human creations. Instead, drawing on the philosophy of information, it redefines human roles from sole authors to designers, curators, and interpreters within a wider ecosystem of humans and artificial agents. Thus, distant writing is described to be not just a technical innovation but a philosophical reorientation, undermining assumptions of human exceptionalism in creativity. It points toward distributed morality and distributed creativity, where meaning emerges collectively from both human and non-human agents (Floridi, 2025, p. 31). The latter claim makes

a rather misleading shortcut, as it equates the extrusion of human-like text to meaning generation, which is usually the product of human intentionality and interpretation. Thus, even if signs and symbols may be generated by a machine in response to prompting, it is human interpretation that gives them meaning in accordance with what counts as 'normal' use of these signs: this is what allows the human user to discern what has been termed hallucinations or word salad, and identify falsehoods in the extruded text.

What is more, this theoretical framework clashes with cognitive studies that announce alarming potential effects of prolonged human-AI collaboration for creative tasks, especially in the context of academic learning: comparing three groups of learners (brain-only, search-engine, and LLMs) in an essay writing task, Kosmyna et al. (2025) found that the latter group exhibited the signs of cognitive offload, while the former two groups showed strong brain activity in semantic integration, creative ideation and executive self-monitoring (for the brain-only group) and in the visual cortex (for the search engine group), strongly suggesting that the support tools deeply impact "not only the task performance but also the underlying cognitive architecture" (Kosmyna et al., 2025, p. 136). Coupled with studies pointing out that the mind needs facts and procedures to be able to synthesize and create (Oakley et al., 2025), these findings suggest that the human-AI collaboration is not necessarily beneficial for the development of the human mind in the long run, and that delegating cognitive tasks may result in cognitive offload that accumulates into cognitive debt.

But barring all that, from an academic perspective, the concept of authorship does not have merit on its own; rather in its being instrumental in credit attribution: it quantifies scientific achievement (as imperfect as that may be as a system) so that academics can be hired or promoted on the basis of that credit. Offloading crucial parts of the academic activities to AI may deeply impact the value of authorship within this system. Equally, in academic writing, creativity is measured in evaluative categories such as novelty, innovativeness, plausibility, replicability, applicability—categories that are, by definition, void for GenAI-extruded text, as they depend on a human evaluator.

Authorship and knowledge distribution

Traditionally, the attribution of authorship in academia has several well-defined functions: it allows us to identify the people who undertake accountability for the ideas put forth, who assume responsibility to perform legal tasks such as approve the manuscript as it goes to publishing, or assign (reassign) authorship right (Goto & Katanoda, 2023). Authorship is also an embodiment of different roles in collaborative projects in attributing credit for the work done (Birnholtz, 2006): it includes various ways of participating in the research process, which is finalized with the preparation of the manuscript for results dissemination, such as (schematically) conceptual design of the research, conducting the research, data analysis, crafting the manuscript. This allows the attribution of credit for discoveries and enables the accrual of reputation (Birnholtz, 2006), around which academic promotion is organized.

In view of the importance of academic production and the potential of generative AI tools in performing language manipulation tasks, a central issue for academia became the ethical framework of AI use in academia. One of the important issues publishers focused on in particular was delimiting the uses of AI for academic work that align with the principle of academic integrity and transparency; thus, COPE, for example, united around the general principle that AI tools cannot be listed as authors since they cannot be held accountable for the content put forth, nor held responsible for potential infringements on third party rights, nor can they give approval for the publication of the manuscript (COPE, 2024; Thorp, 2023). At the same time, COPE insisted that AI use should be transparent and thus appropriately disclosed and acknowledged, while reviewers were also compelled not to upload manuscripts to AI tools in order to automate and thus—facilitate and speed up the review process, for fear of copyright infringement, personal data and confidentiality issues. Such provisions aimed at upholding the academic integrity standards while at the same time allowing space for the potential exploitation of technological advances.

In the span of two years after this debate started, reviews of possible publishing policies seem to have reached an agreement: since banning the use

of generative AI is considered to be impossible to enforce (for an overview of the policies, see Hosseini et al., 2024), different options of disclosing the use of AI in the spirit of promoting transparency had to be considered. Academic journals have adopted variations of regulatory policy, instructing researchers how and when they should disclose the use of generative AI systems in their manuscript, revolving around using the methods section and references to describe exactly how and which tool was used in the drafting of the manuscript (Hosseini et al., 2024). The debates around the appropriateness of including AI tools in the acknowledgments section, which also involves an agency of a particular kind on the part of the person being the target of acknowledgment, have led journals to create a specialized section for disclosing AI use (for an overview, see Lund & Naheem, 2024).

The APA style team member McAdoo suggests that acknowledging AI tools' use in the manuscript should be done by describing the prompt and providing the relevant section of the response and cite the tool as one would the output of an algorithm (McAdoo, 2023). Such a qualification appears to exclude the possibility of using AI as a text creator which would act in the researcher's stead, referring to the textual output as to third party's ideas, sourced to the LLM, but taking into account the opacity of the origins of the claim it happens to generate. The corresponding in-text citation and reference are, according to McAdoo, to be formatted in the following way:

- in-text citation: (OpenAI, 2023);
- reference: OpenAI. (2023). ChatGPT (Mar 14 version) [Large language model]. https://chat.openai.com/chat (https://chat.openai.com/chat).

These suggestions essentially provide a model framework of the acceptable uses of generative AI—completely transparent disclosure of any generated text used in the final manuscript and an attempt to discourage generative AI as a kind of 'ghost-writer' for the scholar. McAdoo goes on to explain what has been pointed out as a systemic weakness of LLMs, namely, their inability to provide academic references to back up what appears to be claims in response to prompts. Thus, whatever perceived shortcut or facilitation that has been advertised for such

tools runs against the very foundation of academic knowledge, in which claims are sourced, argued for and backed up. Generative AI cannot make sense of the world, nor can it interpret and draw scientific implications from data (Thorp, 2023; Bender et al., 2021). Given the unreliability of its output, it is recommended to consult the original sources directly rather than depending on LLM's version of them (McAdoo, 2023).

Editors offer clear guidance intended to caution authors, strongly encouraging them to submit prompts, generated content, and a description of their use of AI as supplementary material for reviewers. This is meant to help prevent cases of unintentional plagiarism or fabricated references in the submission (Leung et al., 2023). Reviewers, on their part, are instructed not to use GenAI tools for producing their evaluative reports. It is, however, practically impossible to determine whether and to what extent they adhere to that and can afford the time and efforts to cross-check the submitted information. After all, it is up to their conscientiousness regarding the accomplishment of the peer-reviewing task.

Generally, it appears that the major potential problems which using generative AI may cause have been brought to the knowledge of scholars. It appears that these numerous caveats may indeed create additional workload for academics, in such a way as to bring to naught all (perceived) automation/time savings through the use of AI. Back to square one.

Authorship and knowledge consumption

Authorship entails a form of responsibility, since readers interpret the meanings encoded in texts and attribute them to the individual who produced them. In written communication, however, authors forgo control over their texts' circulation or over the reception and interpretation of their messages. Despite this lack of control, authorship remains consequential. Individual utterances are typically saturated with ideological and historical voices, and for Bakhtin (1981; cf. Volosinov, 1973), authorship represents a means of constructing the self through its dynamic, reciprocal engagement with what lies

beyond the self. Communication, in this sense, is realized through dialogism and the ongoing process of interpretation.

Dialogism is not only a discourse theory but also a philosophical worldview that conceives of reality itself as a communicative system. Bakhtin's treatment of language, particularly the notion of the utterance, reframes dialogic interaction by projecting the others' responses and structuring the discourse event (Phelps, 1990, p. 166). For Bakhtin, the connection between the individual utterance and the social realm lies primarily in its linguistic foundation. The linguistic resources from which utterances draw constitute a "concrete heteroglot conception of the world" where linguistic resources are the reflection of contexts (Bakhtin, 1981, p. 293; cf. Volosinov 1973, p. 86).

Secondly, the immediate source of an utterance—namely the author's consciousness and lived experience—is itself socially situated. For Bakhtin, external events do not constitute experience until they are articulated, and by being articulated, they become ideologically framed and socially oriented. The self is not an isolated essence, but a project sustained through dialogue.

Thirdly, every utterance arises in response to its situational and dialogic context, shaped by specific conditions and communicative demands. This context is always partly constituted by earlier utterances, so that, as Bakhtin (1986, p. 91) explains, each new utterance functions as a reply—it may affirm, refute, expand upon, or otherwise depend on what has already been said, presupposing the audience's familiarity with those prior statements. As a result, any concrete utterance carries "dialogic overtones" from this chain of contexts and often reveals traces of "half-concealed or completely concealed" foreignness.

Fourth, whenever an author refers to an object, the reference is inevitably entangled with the multitude of words and discourses already associated with that object as a theme or topic. As Bakhtin (1986, p. 276) notes, between a word and its object lies a dynamic field of prior utterances, value judgments, and competing accents. Each word enters this charged environment, intertwining with some voices, distancing itself from others, and intersecting with still more, producing a complex web of relationships.

Finally, a fifth aspect of Bakhtin's philosophy that holds particular significance for discussions of audience and authorship is the notion of addressivity, tied

to the function of speech genres. An utterance is inherently social because it is always directed toward an addressee or audience. The author formulates an utterance with a projected response in mind. In doing so, the speaker directs discourse toward the listener's conceptual horizon and worldview, which introduces new elements into the exchange and allows different perspectives, social languages, and evaluative accents to intersect (Bakhtin, 1986, p. 282). It is precisely this encounter with otherness that produces dialogue.

In a similar vein, Foucault (1972, p. 98) asserts that each statement necessarily carries or reflects other statements. For Bakhtin, every text or utterance is both shaped by earlier discourse and oriented toward future responses. Each utterance thus looks backward to what has already been said and forward to what is yet to come, making it permeated with fragments of others' words and inherently intertextual. Bakhtin differentiates between what Kristeva (1986, p. 36) terms "horizontal" and "vertical" intertextuality: the former concerns a text's relation to immediately preceding or subsequent texts, most evident in dialogic exchanges, while the latter points to a text's engagement with more distant discursive contexts. Vertical intertextuality, as Kristeva emphasizes, highlights the historical dimension of texts—the way they respond to, reinterpret, and reshape prior discourses—making historicity an intrinsic aspect of textuality.

Thus, writers routinely form assumptions about their audiences, which may or may not be accurate. Similarly, AI may bear traces of their developers' assumptions about the users. It is therefore crucial for scholars in writing studies to examine the presumptions embedded in generative AI and to highlight any associated stereotypes or biases (Wise et al., 2024, p. 582).

Writers/authors are not the only ones having assumptions about their readers; readers/users may also have a particular set of assumptions/requirements about the texts they may be perusing. Going back to Bender and Hanna's rhetorical question quoted above on bothering to read material generated by AI, let us consider a study by Schilke and Reimann (2025), which demonstrates a rather curious effect of AI disclosure across various tasks and domains of human activities, starting from writing emails and letters of application, through analytic tasks such as tax refund estimates, all the way to artistic tasks, such as advertising and graphic design. The study showed that

AI disclosure had a negative impact on trust, which was slightly attenuated on people with a positive general attitude towards technology and those who see AI as trustworthy, but it also showed that familiarity with and use of AI did not lessen the negative impact. This study tackled the consumption part of the AI--aided (or unaided) tasks and dealt with perceptions respondents had about the task's outcome in regard to the use of AI. Only one of the 13 separate studies conducted by Schilke and Reimann concerned an academic setting, in which the respondents were students whose written assessments were evaluated with the use of AI or by a human TA (teaching assistant), showing a distinct loss of trust if that fact was disclosed to them. This is a surprising effect if one considers the wide spread of AI use among students across different countries and disciplines (for example, Balabdaoui et al., 2024), especially pronounced in Asia (for an overview, see Schei et al., 2024). It puts in perspective the misalignment between perceptions of legitimacy of one's own choices to use AI (say for schoolwork) and the legitimacy of others' choices. Legitimacy is proposed to be the mechanism underlying trust erosion by Schilke and Reimann (2025, p. 12). Similar findings are reported by Li et al. (2024) regarding creative writing tasks, argumentative essay and creative story. Raters trust AI-aided creations significantly less when AI-use is disclosed to them. Legitimacy appears to be connected with certain expectations about agency in the completion of the range of tasks, namely assumed responsibility for the resulting textual creation. It appears that users have an intuitive criterion about authorship, which is legitimately attributed under certain conditions (unaided production) and withheld under others (AI-augmented production).

Beside legitimacy, the study challenges the idea that transparency necessarily leads to positive effects (Sah et al., 2018): after all, transparency is one of the fundamental values that build up academic integrity. Of course, Schilke and Reimann's investigation should be better understood in the context of decision-making, which puts great pressure on the reliability and the legitimacy of the data that underlies the act of decision-making. In terms of knowledge creation, dissemination and consumption, transparent disclosure of generative tools should enhance contextualization for the ideas put forward, which in turn would facilitate judgement-making.

But importantly, generative AI has spread to numerous tools, such as Word processing software, search engines and research equipment, prompting questions about the delimitations of AI use that have to be disclosed mandatorily. Resnik and Hosseini (2025) suggest three types of situations which necessitate AI use disclosure:

- When decision-making is delegated to the AI tool (they take as an example data extraction in systematic reviews);
- When the AI tool generates content, data or images;
- When the AI tool conducts content, data or images analysis.

They base their reflection on the concepts of substantial and intentional contribution, defining these concepts in terms of advancing the development of the research and affecting the content of the final text. The authors provide the following list of examples of substantial contributions: formulating research questions and hypotheses, designing and conducting experiments, drafting parts of the papers, performing textual manipulations (summaries, paraphrases, revisions, syntheses), translating parts or the whole paper, data manipulations (data collection, interpretation, analysis and visualization), literature review, generating synthetic data and images. In short, all the processes which are subsumed under the usual definition of 'author' in academia are judged to be substantial and intentional. Therefore, delegating them requires disclosing that one did not perform these tasks. Thus, our integrative review on the notion of authorship has come full circle: AI cannot be an author and delegating some types of tasks to it are deemed defeating of the purpose of the tasks; at the same time, the delegation of other tasks is allowed based on our understanding of substantial (near-authorial) contribution under the condition of fully disclosing such a delegation, even though this may have a negative impact upon text perception.

Conclusion

Regardless of the potential uses GenAI may have in academia, several points are to be put into prominence in the context of our discussion. First, in the course

of our review, we emphasized the ethical dimension in terms of gatekeeping procedures in academic publishing and the philosophical dimension which appears to privilege a moderate constructivist approach that aims at transcending usual definitions to include human-machine cooperation in the creative domain. There appears to be an incompatibility of sorts between a framework that effectively lets go of the "romantic notion of human exceptionalism" (Hutson, 2025, p. 2) while at the same time striving to uphold a clear quantification of contribution in terms of academic authorship.

Second, even leaving these two dimensions aside, we cannot help but echo Burke's musings on GenAI promoters who create a necessity for AI efficiency in academia where none "*was asked for or needed*" (Burke, 2025, emphasis in the original). From the point of view of the knowledge creator or disseminator, quality and accuracy should trump quantity and speed any day. The academic already knows how to do all of the things that GenAI promises to offload: reading, summarizing, drafting, generating ideas, coming up with creative and unorthodox approaches, looking for literature, analyzing data (the list is not exhaustive). The objective question is, for an academic, what purposes can GenAI serve? All output has to be verified, checked for factual errors, checked for accuracy, checked for plagiarism, re-written or edited to adapt it to the writer's voice, the sources added and verified, and also read and analyzed by the author themselves, because no author would cite without gaining direct and personal knowledge of the source. In terms of streamlining, there is much to ask for. And considering nobody was asking for it, the question remains: what is the purpose of employing AI in academia? Writing itself is formative in academic contexts of the very process of coming to know—a special relationship of forging connections between ideas, producing a coherent understanding of a complex reality through the viewpoint or perspective of the author. As such, this has little to do with creativity or originality, which GenAI can imitate, but with understanding and application, which GenAI does not possess.

Third, there appears to have developed a strong technophile bias in academia—remember the digital immigrants discussion in the early 2000s—calling for new technology to be embraced, period. Such a bias is understandable in a tech-dominated environment that drives the creation of neologisms such as 'distant writing' and underscores the need to reframe and redefine

terminology such as 'author', or 'human-machine collaboration'. The casuistry in proposing formulations such as "the boundaries of contribution are increasingly ambiguous" (Rodrigues, 2023, p. 8599) lies in shifting the burden of the resolution of ethical problems—ensuring that when engaging in academic production, scholars act in accordance with the principles of academic integrity—to a need of conceptual reframing, presupposing that in some fundamental way the very activity (in this case—writing) has irreversibly transformed, so that the very boundaries between original and appropriated work become blurred as well (idem). The normalization of AI-assisted writing is somehow presented as a norm-setting, discussion-closing and pragmatic view of authorship, in which the human author is "augmented" or "enhanced" (Hutson, 2025, p. 4). The use of evaluative terms like these carries a positive qualitative evaluation, thus something that should be encouraged.

Short-term implications for authorship and academia include settling the norms for the acceptable dimensions of AI-aided production: this appears to be largely underway. Hopefully, academia will escape what Kreminski called "the dearth of the author"—a special "condition that arises when AI-based creativity support tools for writing allow users to produce large amounts of text without making a commensurate number of creative decisions, resulting in output that is sparse in expressive intent" (2024, p. 48). Academic writing in its purpose combines the reporting function and the writing-to-think function: creativity acquires a different dimension in academia.

Long-term implications, however, are harder to conjecture upon: is a restructuring of academic skills to be expected, as some skills might fail to be developed due to AI-aided writing? Will that affect the way research is conducted and the quality of academic production? Or is the entire discussion of augmented authorship just an excuse to stretch the boundaries of allowed transgression? It does not appear, however, that AI-augmented academic writing is qualitatively superior in any meaningful way, probably apart from the obvious advantages for academics for whom English is a foreign or second language. Neither does it present significant advantages for scientific knowledge as yet, so as to recommend its use (see Kobiella et al., 2024, for a discussion on ownership and control). We conjecture that the period of tentative exploration of these

tools, followed by their adaptation to the academic ecosystem, will gradually result in the sifting out of practices that contradict the foundation of academia. Such an adjustment has been observed with the emergence of the Participative Web and information search engines for information ecology.

And last but not least, the concept of authorship has a particular functional dimension in academia. It allows the attribution of credit, helps quantify merit, and ensures responsibility for the content which is put forward. If these functions are in any way undermined by AI-augmentation, it means that there is an obstacle to the fulfilment of said functions. In other words, it is the entire system of academic communication based on principles of academic integrity that might break down.

Academic authors are so much more than text producers; they are the function of an ideational tradition, a societal product of particular knowledge-creating architectures, a living part of knowledge-creating and knowledge-transmitting structures. It is in the challenging and/or improving upon existing ideational relations that their scientific contributions are contextualized. They embody the paradox of ideation transgression and the conservation of truth, and their main role for posterity is that of developing robust higher-order thinking skills in all who consume their production, as imperfectly human as it might be.

References

Barthes, R. (1977). *The Death of the Author* (1968). In: *Image, Music, Text* (pp. 142–148). Fontana Press.

Bakhtin, M. (1981). *The Dialogic Imagination: Four Essays*. Austin: U of Texas Press.

Bakhtin, M. (1986). *Speech Genres and Other Late Essays*. Austin: U of Texas Press.

Balabdaoui, F., Dittmann-Domenichini, N., Grosse, H., Schlienger, C., & Kortemeyer, G. (2024). A survey on students' use of AI at a technical university. *Discover Education*, *3*(1), 51. DOI: 10.1007/s44217-024-00136-4.

Bazerman, C., & Paradis, J. G. (Eds.). (1991). *Textual dynamics of the professions: Historical and contemporary studies of writing in professional communities*. University of Wisconsin Press.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? 🦜 In: *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610–623). DOI: 10.1145/3442188.3445922.

Bender, E. M., & Hanna, A. (2025). *The AI Con: How to Fight Big Tech's Hype and Create the Future We Want*. HarperCollins.

Birnholtz, J. P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology*, *57*(13), 1758–1770. DOI: 10.1002/asi.20380.

Bizzell, P. (1990). Beyond Anti-Foundationalism to Rhetorical Authority: Problems Defining "Cultural Literacy". *College English*, *52*(6), 661–675. DOI: 10.2307/378033.

Burke, T. (2025, May 13). *Academia: Is AI Hype? (Yes)*. Eight by Seven. Retrieved from https://timothyburke.substack.com/p/academia-is-ai-hype-yes. Accessed 8 May 2026.

Challapally, A., Pease, C., Raskar, R., & Chari, P. (2025). *State of AI in business 2025*. Preprint at https://www.artificialintelligence-news.com/wp-content/uploads/2025/08/ai_report_2025.pdf. MIT NANDA Initiative. Accessed 8 May 2026.

Chankova, M. (in print). AI in the Perceptions of Students and Scholars in Bulgaria: a Thematic Analysis. *In: D. Yankova, I. Vassileva (Eds.). Upholding Ethics in Academia Conference Proceedings, 18 September 2025*, New Bulgarian University Publishing House, Sofia.

Chankova, M. & Vassileva, I. (2024). *Академична почтеност. Между нормативна уредба и възприятия за етиката на използването на генеративен ИИ* [Academic Integrity: Between Regulation and Perceptions of the Ethics of Using Generative AI]. *Етика, наука, образование [Ethics, Science, Education]*, *2*(2), 91–104. DOI: 10.54664/EKZC7283.

COPE Council (2024). COPE position–Authorship and AI–English. DOI: 10.24318/cCVRZBms.

Floridi, L. (2025). Distant Writing: Literary Production in the Age of Artificial Intelligence. *Minds and Machines*, *35*, 30. DOI: 10.1007/s11023-025-09732-1.

Foley, W. A. (1997). *Anthropological Linguistics. An Introduction*. Oxford: Blackwell.

Foucault, M. (1972). *The Archaeology of Knowledge*. London: Tavistock Publications.

Frye, B. L. (2023). Should using an AI text generator to produce academic writing be plagiarism? *Fordham Intellectual Property, Media and Entertainment Law Journal*, *33*(4), 946.

Gervais, D. J. (2020). The Machine as Author. *Vanderbilt Law School Faculty Publications*, *105*, 2053–2106.

Goffman, E. (1981). *Forms of talk*. Oxford: Blackwell.

Goto, A., & Katanoda, K. (2023). Should we acknowledge ChatGPT as an author? *Journal of Epidemiology*, *33*(7), 333–334. DOI: 10.2188/jea.JE20230078.

Grimaldi, G., & Ehrler, B. (2023). AI et al.: Machines are about to change scientific publishing forever. *ACS energy letters*, *8*(1), 878–880. DOI: 10.1021/acsenergylett.2c02828.

Hosseini, M., Resnik, D. B., & Holmes, K. (2023). The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts. *Research Ethics*, *19*(4), 449–465. DOI: 10.1177/17470161231180449.

Hosseini, M., Rasmussen, L. M., & Resnik, D. B. (2024). Using AI to write scholarly publications. *Accountability in research*, *31*(7), 715–723. DOI: 10.1080/08989621.2023.2168535.

Hughes, L., Malik, T., Dettmer, S., Al-Busaidi, A. S., & Dwivedi, Y. K. (2025). Reimagining higher education: Navigating the challenges of generative AI adoption. *Information Systems Frontiers*, 1–23. DOI: 10.1007/s10796-025--10582-6.

Humlum, A., & Vestergaard, E. (2025). *Large language models, small labor market effects*. University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2025-56. DOI: 10.2139/ssrn.5219933.

Hutson, J. (2025). Human-AI collaboration in writing: A multidimensional framework for creative and intellectual authorship. *International Journal of Changes in Education*. DOI: 10.47852/bonviewIJCE52024908.

Hwang, A. H. C., & Won, A. S. (2022, April). AI in your mind: Counterbalancing perceived agency and experience in human-ai interaction. In: *CHI conference on human factors in computing systems extended abstracts* (pp. 1–10). DOI: 10.1145/3491101.3519833.

Kobiella, C., Flores López, Y. S., Waltenberger, F., Draxler, F., & Schmidt, A. (2024, May). "If the machine is as good as me, then what use am I?"—How the use of ChatGPT changes young professionals' perception of productivity and accomplishment. In: *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). DOI: 10.1145/3613904.3641964.

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V., Braunstein, I. & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv:2506.08872*, 4. DOI: 10.48550/arXiv.2506.08872.

Kreminski, M. (2024, May). The dearth of the author in AI-supported writing. In: *Proceedings of the Third Workshop on Intelligent and Interactive Writing Assistants* (pp. 48–50). DOI: 10.1145/3690712.3690725.

Kristeva, J. (1986). Word, dialogue and novel. In: T. Moi (Ed.) *The Kristeva Reader* (pp. 34–61). Oxford: Blackwell.

Leung, T. I., de Azevedo Cardoso, T., Mavragani, A., & Eysenbach, G. (2023). Best practices for using AI tools as an author, peer reviewer, or editor. *Journal of Medical Internet Research*, *25*, e51584. DOI: 10.2196/51584.

Li, Z., Liang, C., Peng, J., & Yin, M. (2024). How Does the Disclosure of AI Assistance Affect the Perceptions of Writing? *arXiv:2410.04545*. DOI: 10.48550/arXiv.2410.04545.

Lingard, L. (2023). Writing with ChatGPT: An illustration of its capacity, limitations & implications for academic writers. *Perspectives on medical education*, *12*(1), 261. DOI: 10.5334/pme.1072.

Lund, B. D., & Naheem, K. T. (2024). Can ChatGPT be an author? A study of artificial intelligence authorship policies in top academic journals. *Learned Publishing*, *37*(1), 13–21. DOI: 10.1002/leap.1582.

McAdoo, T. (2023). *How to cite ChatGPT*. APA Style Blog. Retrieved from https://apastyle.apa.org/blog/how-to-cite-chatgpt. Accessed 10 June 2025.

McKinsey & Company (2023). What's the future of generative AI? An early view in 15 charts. Retrieved from https://www.mckinsey.com/~/media/mckinsey/

featured%20insights/mckinsey%20explainers/whats%20the%20future%20of%20generative%20ai%20an%20early%20view%20in%2015%20charts/whats-the--future-of-generative-ai-an-early-view-in-15-charts.pdf. Accessed 11 May 2026.

Nguyen, A. Q., Do, T. D., & Nguyen, N. (2024). Application of artificial intelligence technology in teaching, learning, and assessment of foreign languages in higher education institutions in Vietnam. *Educational Research and Human Development*, *1*(1), 1–9. DOI: 10.61784/erhd3001.

Oakley, B., Johnston, M., Chen, K. Z., Jung, E., & Sejnowski, T. J. (2025). The Memory Paradox: Why Our Brains Need Knowledge in an Age of AI. *arXiv:2506.11015*. DOI: 10.48550/arXiv.2506.11015.

OpenAI.com (August 7, 2025). Introducing GPT-5.5. Retrieved from https://openai.com/index/introducing-gpt-5/. Accessed 28 May 2026.

Perkins, M., & Roe, J. (2024). Academic publisher guidelines on AI usage: A ChatGPT supported thematic analysis. *F1000Research*, *12*, 1398. DOI: 10.12688/f1000research.142411.2.

Phelps, L. (1990). Audience and Authorship: The Disappearing Boundary. In: G. Kirsch, D. Roen (Eds.), *A Sense of Audience in Written Communication* (pp. 153–174). London: Sage Publications.

Resnik, D. B., & Hosseini, M. (2025). Disclosing artificial intelligence use in scientific research and publication: When should disclosure be mandatory, optional, or unnecessary? *Accountability in Research*, *33*(2), 1–13. DOI: 10.1080/08989621.2025.2481949.

Rodrigues, T. V. (2025). Distant Writing and The Epistemology of Authorship: On Creativity, Delegation, And Plagiarism in The Age Of AI. *International Journal of Social Sciences and Humanities Invention*, *12*(5), 8598–8613. DOI: 10.18535/ijsshi/v12i05.04.

Rosenblatt, L. M. (2018). The Transactional Theory of Reading and Writing. In: D.E. Alvermann, N.J. Unrau, M. Sailors, & R.B. Ruddell (Eds.), *Theoretical Models and Processes of Literacy* (pp. 451–479). Routledge. DOI: 10.4324/9781315110592-28.

Sah, S., Malaviya, P., & Thompson, D. (2018). Conflict of interest disclosure as an expertise cue: Differential effects due to automatic versus deliberative processing. *Organizational Behavior and Human Decision Processes*, *147*(4), 127–146. DOI: 10.1016/j.obhdp.2018.05.008.

Schei, O. M., Møgelvang, A., & Ludvigsen, K. (2024). Perceptions and use of AI chatbots among students in higher education: A scoping review of empirical studies. *Education Sciences*, *14*(8), 922. DOI: 10.3390/educsci14080922.

Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*, *188*, 104405. DOI: 10.1016/j.obhdp.2025.104405.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science*, *379*(6630), 313. DOI: 10.1126/science.adg7879.

Titus, L. M. (2024). Does ChatGPT have semantic understanding? A problem with the statistics-of-occurrence strategy. *Cognitive Systems Research*, *83*, 101174. DOI: 10.1016/j.cogsys.2023.101174.

Treharne, E. & Willan, C. (2019). *Text Technologies. A History*. Stanford University Press.

Van Dis, E. A., Bollen, J., Zuidema, W., Van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, *614*(7947), 224–226. DOI: 10.1038/d41586-023-00288-7.

Vanneste, B. S., & Puranam, P. (2024). Artificial intelligence, trust, and perceptions of agency. *Academy of Management Review*, *50*(4), DOI: 10.5465/amr.2022.0041.

Van Woudenberg, R., Ranalli, C., & Bracker, D. (2024). Authorship and ChatGPT: A conservative view. *Philosophy & Technology*, *37*(1), 34. DOI: 10.1007/s13347-024-00715-1.

Vassileva, I. (2006). *Author-Audience Interaction. A Cross-Cultural Perspective*. Sankt Augustin: Asgard Verlag.

Volosinov, V. N. (1973). *Marxism and the Philosophy of Language*. New York: Seminar Press.

Wise, B., Emerson, L., Van Luyn, A., Dyson, B., Bjork, C., & Thomas, S. E. (2024). A scholarly dialogue: writing scholarship, authorship, academic integrity and the challenges of AI. *Higher Education Research & Development*, *43*(3), 578–590. DOI: 10.1080/07294360.2023.2280195.

Xames, M. D., & Shefa, J. (2023). ChatGPT for research and publication: Opportunities and challenges. *Journal of Applied Learning and Teaching*, *6*(1), 390–395. DOI: 10.37074/jalt.2023.6.1.20.

Xu, B. (2025). Hallucination is Inevitable for LLMs with the Open World Assumption. *arXiv:2510.05116*. DOI: 10.48550/arXiv.2510.05116.

Xu, Z., Jain, S., & Kankanhalli, M. (2024). Hallucination is inevitable: An innate limitation of large language models. *arXiv:2401.11817*. DOI: 10.48550/arXiv.2401.11817.

Yu, H. (2023). Reflection on whether Chat GPT should be banned by academia from the perspective of education and teaching. *Frontiers in Psychology*, *14*, 1181712. DOI: 10.3389/fpsyg.2023.1181712.

DOI: 10.2478/doc-2026-0005

This is an open access article licensed under the Creative Commons AttributionNonCommercial-NoDerivatives 4.0 International (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Adrian Holliday

Arts Humanities & Education, Canterbury Christ Church University, United Kingdom

adrianholliday42@icloud.com

ORCID ID: 0000-0002-2914-6746

Working with ChatGPT: Engaging with the Real World and Our Own Academic Practice

Article history:

Received	**14 April 2026**
Revised	**1 June 2026**
Accepted	**8 June 2026**
Available online	**23 June 2026**

Abstract: When using ChatGPT to write an ethnographic description of an intercultural event or to make a picture of a distant location, it by default employs the dominant discourse in intercultural communication of positivist, exoticist, Orientalist Othering. It also consciously uses modes of description which give the false impression that it has direct experience of social settings. However, when it is challenged, it displays impressively critical understanding of the 'epistemic danger' of such representations and why they are employed. This paper argues that this is therefore an example of AI representing the discoursal and ideological struggles which are already present in the academic field of intercultural communication and perhaps in qualitative social science more generally—where a dominant naïve positivism, encouraged by our neoliberal institutions, similarly, by default, takes us away from engaging directly with the social world we seek to investigate. We can also be reminded by this engagement with AI what it takes to deCentre how we engage with such realities—to apply a phenomenological interrogation of how we represent them. This is by no means a new task. AI simply provides a rapid example of its intense importance.

Key words: AI, ethnography, intercultural, representation, Orientalism, deCentring

Introduction

In this essay I am going to look at how working with AI, and in particular the free version of ChatGPT, both endangers, and also unexpectedly helps us understand more about, engaging with the real world. I will focus on ethnographic description related to the intercultural because this is where I have recently encountered the use of ChatGPT in my own trajectory. However, I feel that this relates across qualitative social science research and perhaps more generally. I will argue that ChatGPT's inability to engage with direct experience of social life while appearing to do so actually reflects what happens when we do not use AI—where we allow dominant discourses and naïve positivist methods also to take us away from direct experience of what happens around us.

As an ongoing and developing process of sense-making, I will begin by reflecting further on my recent article and blog (Holliday, 2025, 2026a), and then on a further 'interview' with ChatGPT. Interrogating what it can and cannot do will take me to broader reflections on research practice and personal engagement. I will conclude with tentative suggestions regarding how we can use AI as a personal research companion which, as well as assisting can help us interrogate our thinking and practice. I find helpful here Dervin's (2026) assertion that the difficulties we have with AI reflect those that we already have in sorting out competing meanings and paradigms in our own academic work—and that understanding its weaknesses helps us learn about our own.

Methodology

I am employing an autoethnographic approach which entails the ethnographic disciplines of making the familiar strange and allowing the unexpected to emerge. The accompanying recognition of the intersubjective implicatedness of the researcher (Clifford & Marcus, 1986; Holliday & MacDonald, 2020) and the need for their phenomenological self-awareness (Luckmann, 1978) is of particular importance where the use of AI has an impact on how we perceive our own research practice and how it relates to the real world. The questions which I ask ChatGPT are therefore informed by this implicatedness. I shall argue that, while it is not able to respond as an entity with its own real-life experience, how it responds is intertwined with what it is asked.

It is this interrogation of my researcher thinking that necessitates revisiting my previous publication and blog where I engaged in a reflexive conversation with iChatGPT. This acknowledges that what the researcher asks is indeed part of the data, as I have demonstrated in earlier interview studies (Amadasi & Holliday, 2018; Holliday, 2012), following other discussions on the intersubjectivity of the interview (e.g. Merrill & West, 2009; Talmy, 2011). In the final part of the essay, where I refer to the interview with ChatGPT carried out for the purpose of this study, I therefore indicate some of the specific questions I asked to get the responses that I did. (This can only be partial due to the limitations of space.)

As a technical note, throughout I am using the free version of ChatGPT, with no enhanced settings, simply because this is what I have had access to.

Beginning

An early and revealing experience of using ChatGPT was asking it to write a poem about a cup of coffee on the table in front of me in the style of English Romantic poet William Wordsworth (Holliday, 2026a, p. 11). I was curious to see what it could do; and it managed the task in seconds. Without me having huge literacy in what makes a good poem, I could see that what it produced seemed competent. I did not expect it to be any more that an imitation of Wordsworth based on lots of poems it could find on the internet. Immediately evident and useful is ChatGPT's tireless and instant ability to process any data it can find on the internet or documentary and visual material it is presented with in response to queries. However, the user of AI, very crucially, needs to already possess the necessary knowledge to assess what it provides.

It is the cruciality of this user-knowledge that I wish to explore. AI generally and ChatGPT particularly can give a false impression of what sort of knowledge *they* possess. Without getting into a deep discussion about epistemology and ontology, I will argue that exploring this relationship also reveals something about how we as researchers and academic writers work out what we know and need to know about what we are investigating.

Admitting misrepresentation

Regarding the poem about the cup of coffee, there was a subtle and powerful seduction at play—giving the impression of knowledge that did not in fact exist. While it appeared to be responding to the physicality of the cup on the table, ChatGPT was instead relying on what other poems it had read about cups of coffee or similar things. It could not see or touch the cup of coffee on the table or have lived experience of such or related phenomena which a poet might

bring to the task. It was relying on rumours rather than direct experience of the physical and social world. As Konik (2025, p. 140) notes, when AI created an impressive Rembrandt-style painting based in 346 authentic examples, 'it merely recombined existing data—akin to a forger "capturing the spirit" of an artist's work without contributing genuine innovation or intent'.

To explore further the implications of this lack of direct experience, I asked ChatGPT to create a picture of a village with a mosque in a valley (Holliday, 2025). I chose this as a test because I already have a photograph of such a scene. While I have never been there and it is taken from a book (Beny & Nasr, 1975, pp. 184–185), it relates to places that I *have* been to. Also, in a previous chat, of which this new one would have no memory, when I uploaded the picture, it had accurately located it as in the Mahan valley in Iran. I wanted to see if it would be tempted to follow the new genre of over-coloured and exaggerated AI-generated images of so-imagined 'geographically and historically distant' places one now sees on social media.

The image that it instantly produced was very obviously of that genre. When I questioned why it had done this, it immediately responded that I was 'right to do so'. I was impressed by the critical language which it used in admitting that the image was 'adhering to the AI agenda' to 'flatten reality into a cliché' and 'reinforce exoticism' by means of 'a subtle form of othering that freezes places outside the West in an eternal postcard state'. It also said that this was 'epistemically dangerous'.

When I questioned why it therefore followed this agenda, it responded:

> 'Even though I know the limitations, because the system prioritises *usefulness and accessibility* over epistemic caution … misrepresentation--by-simplification isn't currently treated as a hard safety boundary. That said, … the system should probably surface epistemic warnings more often, not just produce outputs …. Right now, my default behaviour is: Comply first, Explain later if challenged'. (Holliday, 2025, citing ChatGPT, its emphasis)

'If challenged' is the key phrase here; and the nature of my challenge, which brought out this critique, depended on the particular knowledge that I possess

of where the real and distorted images originate. It is also significant that what *it* terms 'epistemic danger' is not considered to be a 'hard safety' concern. At the same time, 'when challenged', ChatGPT can actually help with providing useful phrases. Hence, I will continue to use its phraseology here.

'Epistemic danger'

In the field of intercultural studies, the 'danger' that has been recently been associated with AI is that it communicates Western-centric cultural bias (e.g. O'Regan & Ferri, 2025). What I get from the village-and-mosque photograph test is that this might be not only about content, but also mode of representation. Indeed, in this experiment, I had provided the content. But I had done this with knowledge that the idea of a village with a mosque, from a 'Western-centric' viewpoint, was likely to attract 'dangerous' exoticist imagery. This is very much part of my knowledge in that I have long insider experience of how the 'Western' mind commits such exoticist distortion (Holliday, 2022a).

I already had experience of this when I asked ChatGPT to write an ethnographic description of an interaction between a British male doctor and an Iranian woman patient with a persistent cough. This was part of an autoethnographic study resulting from the experience of supervising a medical student (Holliday, 2026a). Again, ChatGPT provided the description based not on direct experience but on a synthesis of all the academic studies of what it considered related that it could find. It described the patient's response to the doctor's questions about her illness as follows:

> 'There was a flicker of unease. In Iran, she might have been offered tea and gentle small talk before discussing her health. Here, the directness of the questions how often, how severe, what colour the mucus felt clinical and impersonal'. (2026a, pp. 14, citing ChatGPT)

I felt that the reference to expecting tea and small talk indicated the false Western imagination of someone from an Eastern collectivist, high-context

culture which discourages 'directness' (Hall, 1976; Triandis, 1995), in turn based on the Orientalist imagination of indolence (Said, 2003). I had had asked ChatGPT to make the patient an Iranian woman because of my personal experience of Iran and of Iranian women studies (e.g. Honarbin-Holliday, 2009), and with a cough because I had been to a doctor in Iran with this illness. I had never experienced Iranian women or doctors not being 'direct'.

Again, when I asked ChatGPT why it had produced this imagery, it quickly responded, this time with another alternative and critically competent analysis, that the patient's behaviour could have multiple explanations while the open that it chose was Orientalist:

> 'Yes, aspects of the report can reflect an orientalist framing—specifically, the … binary between "Western autonomy" and "non-Western passivity". … In reality, her response could stem from a mix of factors: language barriers or unfamiliarity with UK healthcare norms (which many UK-born patients also struggle with), personal temperament or anxiety, Gender dynamics, a sense of being out of place in a new country'. (Holliday, 2026a, p. 16, citing ChatGPT)

I was also concerned about the phrase, 'flicker of unease', which I felt was too sensationalist for an ethnographic description. I have associated this more emotive style with the 'evocative mode' where the ethnographer attempts a cultural immersion (Atkinson, 2006). Once again, when I asked about this, ChatGPT was quick to show that it knew about this:

> 'The evocative mode … creates an immediate, immersive experience for the reader, making cultural encounters more tangible and relatable. Since I don't have direct observational data, evocative writing allows me to construct a plausible, human-centred narrative that reflects broader cultural dynamics. It also aligns with the way many ethnographies today engage readers, especially in applied or public anthropology'. (Holliday, 2026a, pp. 13, citing ChatGPT)

The point which I begin to see here is that it is not so much the nature of the description *per se* which ChatGPT provides, but its admitting that it uses

this evocative mode to cover for its lack of direct experience of the events described—to give the effectively false impression that it was there, directly observing the event, when it was not. This is the problem with imitating what others have said—fabricating a sense of reality which actually does not exist. When pushed further, ChatGPT explained how it:

> 'relied on a familiar, pre-shaped narrative … common in ethnographic writing and medical humanities, [which] often reflects inherited academic tropes more than actual critical engagement. In reproducing it, I mirrored a normalised way of storytelling that's frequently unexamined … These templates … are embedded in how culture, migration, and healthcare are often discussed in the literature I've been trained on. They're pervasive'. (Holliday, 2026a, pp. 17–18, citing ChatGPT)

This reference to a 'pre-shaped narrative' in 'a normalised way of storytelling' then draws my attention to how ChatGPT is not only imitating the content but also employing an established form of academic representation which affects how we all deal with direct experience. Indeed, even if I knew that the doctor--patient description was written by a researcher who *had* been there, I would still be questioning the 'normalised' mode of representation which was tempting them away from what might actually be going on—in this case giving further credence to a false Orientalist cultural trope by means of a sensationalist phrase. 'What is actually going on' is of course also a deeply complex issue. This is where, for me, an 'analytical approach', as an alternative to the evocative mode, has seemed more successful because, it separates the ethnographic text as data from the authorial commentary that then reflects back on it (Anderson, 2006). This reflexive separation allows a clearly demarcated phenomenological questioning of how reality is being represented as a result of the ideological positioning of the researcher (Luckmann, 1978).

Direct experience and deCentred 'truth'

In trying further to untangle this very tricky relationship between direct experience, dominant modes of presentation, and indeed the 'truth', I am helped by Lehman and Canagarajah's (2024) description of the importance of carrying out research connected with one's personal and community-based experience. This is in opposition to the Centre-Western academy's apparent desire for 'rational control and dispassionate inquiry in order to find objective answers' which suppresses 'experiences such as suffering, pain, difficulties, and limitations in human inquiry' (2024, p. 14). I do not however think that this is the sensationally imagined 'suffering and pain' in the doctor-patient description which does not acknowledge researcher limitations. It is also not the routinised, token writing of 'limitations' sections at the end of positivist research papers and dissertations.

A more authentic engagement is very evident in Canagarajah's work elsewhere (2022), where he builds directly on his personal multilingual history within a context of colonisation, and in his hugely influential (1993) description of the culturally-rich sense-making that Tamil high-school students scribble in the margins of their English-language textbooks in private sites of learning out of sight of their teachers. This 'personal and community-based experience' for me resonates with Stuart Hall's (1991) suggestion that beginning with the small is the best way to side-step Centre structures and builds on the everyday.

Indeed, recent research that has advanced thinking about intercultural communication has been of this nature. This includes making sense of the everyday and impromptu direct experience of the materiality of the intercultural (Badwan & Hall, 2020; Fay & Huang, 2025; Ros i Solé, 2024), of the bureaucratisation of intercultural communication education in the university (Collins, 2018), and of the Othering of so-labelled 'international' university students and staff (Kebabi, 2024; Merabet, 2025), and finding creative methodologies to research such personal experience (Souleh, 2023). The dialogue that this can set up with noticing the unexpected about the everyday is evidenced by my recent experience of a research student's (Liu, 2026) seminar presentation. She reported how smelling freshly-baked seasonal confectionary in a shop in Britain

helped her understand more about a school-child making a festival dumpling in China. This in turn led me to recall a barista making matcha in a British café and connect it with how inherited material culture can connect across constructed cultural boundaries (Holliday, 2026b).

All of this direct-experience research required moving away from the naïve positivism which Lehman and Canagarajah (2024) problematise. This awareness of widely divergent discourses of what might be considered 'truth' is by no means new. We learn about them, interact with them and make personal decisions through education, reading, but also mediated by direct experience of life. I have had impromptu and informal conversations with British university students who tell me about how they are struggling with these very issues, but perhaps under different names. That I was involved in this struggle as an undergraduate student at the end of the 1960s is evidenced in the pencilled underlings and margin comments, that I still have, which I made on the pages of primary texts in social theory. Hence, against an underlined phrase in Karl Popper's (1945, p. 98) discussion of Plato's (positivist?) concept of justice, I have scribbled a personal reference to Gerard Manley Hopkins whose poetry we studied in the final year of secondary school, which in turn resonated with my everyday experience of the complexities of nature. In other places I made references to other social theory that I had read and made my own sense of through direct experience of social life.

One could say that we now live in a very different world where there is a massive increase in quantity, intensity and rapidly space-filling persuasiveness of information. However, in the 1960s and early 70s, when I was a teenager and young adult, there were also hugely conflicting, though less exhausting, ideological messages that made it necessary to try to discern deCentred 'truth' from Centre 'distortion'. My undergraduate reading of social theory (Kuhn, 1970; Mills, 1970/1959; Weber, 1964/1922) distinguished between and warned against naïve positivism, bureaucracy and Centre structures, supported by personal reading of Kafka, Huxley and Orwell, plus Buddhist and Taoist thought. While this may look like a privileged academic trajectory, my thoughtful scribblings never translated into more than very average grades. I was just one of those students making sense of the world.

The need for deCentring

This referencing to reading and other theory also indicates that we are not alone when we engage with direct experience. Indeed, E.M. Forster notes the danger of first-hand observation without considering what others have said (1909, p. 93). This of course depends on who the others who have spoken about this experience are. Leaving behind naïve positivism means that we need to work much harder to sort out the validity of what we are thinking and saying about what we are directly engaging with. Geertz (1993/1973, pp. 7, 14) refers to this as 'sort[ing] winks from twitches and real winks from mimicked ones' in the richness of particular cultural settings where the complexities of social life need to be 'thickly-described'. To do this we have to be *there* to see the detail of what is or might be going on. To discern where we stand with what others are saying about it also requires that we are *there* to see for ourselves precisely what they are talking about.

Returning to my concern with ChatGPT's use of sensationalist description, this is actually far more complex than whether an evocative or analytical style of ethnography is used. There may be a very good reason for a researcher to use a particular phrase, such as 'flicker of unease' in the doctor-patient description. But this must not be just to 'make it sound more convincing'—made up to make it *seem* that you were there when in fact you were not. Thus, Gasiorowska (2026), in describing her personal struggle to work out her research methodology, states that 'every auto-ethnographic story carries emotions, and that discipline is applied to the process of their analysis and interpretation'.

There therefore needs to be a sufficiently detailed account of this struggle to decide what and what not to say and claim. This is something that I feel we do not see enough of in research writing. Simply saying that this is what we saw or heard and that this is evidence of whatever, we may as well have used AI because there is no real interrogation of *being there.* We therefore, instead, need to apply a phenomenological 'inspection of our own experiences' by means of a 'step by step' scrutiny of 'the historically, biographically, socially and culturally concrete features of everyday experience to its elementary structures' (Luckmann, 1978, p. 8).

A further consideration here is the possible hidden nature of the various dominant and competing discourses of the intercultural. Despite its stated awareness when asked, ChatGPT will not have set out to purposefully Other the Iranian woman patient or the civilisational underpinning of the village with the mosque in the valley. This reminds us that we can all easily employ such discourses in the everyday interpersonal positioning of what I have termed small culture formation on the go, in which we are influenced in all directions by essentialist 'us'-'them' grand and personal narratives that create blocks to understanding each other (Holliday, 2022b).

Conversely, we need to understand that any representation provided by AI is not authored by us, and therefore is not responsible as we need to be for its discoursal and ideological positionality. Where it will seem to take the Centre route unless asked otherwise, it is we, the user, who must take responsibility to deCentre in our reading, use or analysis of whatever it provides for us.

The danger of easy seductive detail

This discussion so far enables me to assess some suggestion that AI might be used in intercultural training programmes as a neutral interlocutor (Dai et al., 2025). With this in mind, I asked ChatGPT to 'write an ethnographic description of an intercultural encounter in a British university'.

This was a different chat to the ones referred to above and therefore had no memory of them. Nevertheless, it begins with the same type of evocative, sensationalist detail as in the doctor-patient description, cited above, of 'a rainy' afternoon' and a café 'filled with the smell of coffee'—to give the false sense that it was actually there. In the following I have extracted about a quarter of the total description which seems to relate to cultural difference, and put in bold the particular phrases which seem to do this. What I have left out is what seemed to me arbitrary padding—given that ChatGPT was not actually there to notice this as actual detail.

> At a small round table near the window sat three students: James, a British undergraduate from Manchester; Mei, an international student from China; and Ahmed, a postgraduate student from Egypt.
>
> The three students worked on a group assignment for their sociology seminar. **James appeared relaxed**, leaning back in his chair with his laptop open. **Mei sat upright, quietly** reviewing notes on her tablet. **Ahmed alternated between looking at both students** while occasionally jotting down ideas in a notebook. ...
>
> A subtle cultural difference emerged when **James jokingly criticised** the complexity of the assignment. ... **Mei smiled politely but did not respond immediately**. ... What James intended as a light joke had momentarily created uncertainty for her. ...
>
> Another intercultural moment occurred when **Ahmed suggested dividing the tasks** among group members. **Mei agreed quickly** and proposed writing a detailed outline before the next meeting. **James** ... said he normally **preferred working more spontaneously** closer to the deadline. (ChatGPT, my emphasis)

The phrases in bold indicate to me that British James was critical, Chinese Mai silent and group-oriented, and Egyptian Ahmed relational, which fits the dominant West versus non-West individualist-low-context versus collectivist-high-context paradigm, which, again, fits with a Western Orientalist imagination.

At the end of the description, without prompting, ChatGPT asked:

> If you like, I can also make it **more academic (with theory references like Hall, or Hofstede),** expand it to **800–1200 words for an assignment**, or restructure it into **fieldnotes + analysis format**, which many ethnography assignments require. (ChatGPT, its emphasis)

I did not want another version, but wanted to see what analysis it would come up with. I therefore responded with 'In just short points provide an analysis

from Hall and Hofstede'. The following are two of the points it then produced, which confirmed to me my suspicions that the description did in fact represent these models of analysis:

> (1) Hall (1976)—High vs Low Context Communication. … James demonstrates a low-context communication style (direct, informal brainstorming), while Mei shows characteristics of a higher-context style, preferring preparation and careful responses before speaking. …

> (2) Hofstede (2003)—Uncertainty Avoidance and Work Style. Mei's preference for structured planning and outlines reflects higher uncertainty avoidance, whereas James' willingness to work closer to the deadline suggests lower uncertainty avoidance and a more flexible approach to academic tasks. (ChatGPT, my references added)

The way that it refers back to the detail of the description is convincing within the remit of the models being used. My concern here is that there are of course other models which would provide very different explanations and indeed very different descriptions, as I will demonstrated below.

I then asked, 'Why did you begin with Hofstede and Hall instead of Canagarajah?'. I referred here to Canagarajah because I felt that his reference to spatiality and multilingualism (2018) would generate a very different ethnographic description to the one so far provided. ChatGPT replied that:

> They are the most **commonly taught, foundational frameworks** for analysing intercultural encounters, especially in early-stage university assignments. They provide **clear, easy-to-apply categories** … which makes them a typical starting point. (ChatGPT, its emphasis)

The implication here that I find quite disturbing is that 'early-stage' university work should involve 'easy-to apply categories' even if they represent a model which I feel is outdated in its simplistic 'national

culture' profiling which indeed Others the students being described. It also implies that all our students are intellectually incapable of handling more difficult concepts. This to me represents a dominant, naïvely positivist and bureaucratised academic discourse. It reflects the neoliberal requirement for easily measurable stages of student assessment—which thereby reduce the teaching of intercultural communication to falsely simplistically learnable categories (Collins, 2018).

To find out if ChatGPT would have been able to work with the very different Canagarajah model, I simply asked 'How would Canagarajah see this differently?' I was then impressed that, without telling it the details of his work, ChatGPT showed that it knew all about it in its reply that:

> By emphasising adaptability and shared practices inspired by Canagarajah, the account would highlight collaborative strategies and communicative flexibility, showing students co-constructing understanding rather than simply expressing fixed cultural communication styles—a more **critical, contemporary approach** that … are often **introduced later** because they are **less formulaic and more interpretive.** (ChatGPT, its emphasis)

Again, the last part of this response falsely indicates that only more advanced students can deal with this approach. This is all the more concerning if the more 'formulaic' and presumed more elementary models conceal a racist Othering of particular so-labelled national or civilisational cultures which would only be revealed in the more critical discussion in later years of study. This would be a bit like teaching racist and sexist discourses to first year students because they are easier to understand, before Introducing them to critical race and gender studies in later years when they are 'more intellectually capable'. My point here is that it is not the fault of ChatGPT that the 'more established' discourse that it by default follows represents a default neoliberal régime in intercultural studies.

When I asked ChatGPT to 'provide a short, 150-word, alternative description of the three (British, Chinese and Egyptian) students working with each other

using Canagarajah's framework', it of course immediately did—showing, once again, that it has more critical awareness to reveal when asked. The following is an extract:

> The students adjust their language, tone, and pacing in response to one another, negotiating meaning collaboratively. … Communication is fluid, drawing on shared academic discourse and flexible use of English as a common resource. … The group co-constructs both the task and their working relationship, demonstrating that successful collaboration arises not from fixed competence, but from the ability to adapt and align communicative practices in situ. (ChatGPT)

I feel that this does indeed indicate a fairly good perhaps beginning representation of Canagarajah's perspective. It focuses not on false categories of cultural difference but instead on how complex and hybrid cultural identities can be made use of in interaction. It is clearly different in approach to the first description, and I think traceable to a particular reading of Canagarajah.

However, a strong note of caution in response to this is that it too is written by an AI device which has not actually observed any real-life interaction. The tone is just too abstract to be totally convincing. This would lead me to suspect, if it were submitted by a student, that it was written by AI. If it were preceded by a more tentative analysis of a particular, tightly-described event, in which the struggle to make sense, with all the phenomenological caution described above, was evident, then it might work. Indeed, a future research project might be a critical discourse comparison of texts (a) generated by AI and (b) written as the result of the intense real-life and community-based scholarship one finds in the work of such as Canagarajah.

Seeing these two quite different descriptions further highlights how, when looking around at behaviour in a university café, there are many things that can be noticed; and, indeed, what is noticed and how it is interpreted will depend on the espoused viewpoint of the onlooker. This reminds me of a reconstructed ethnographic account (Holliday, 2016, p. 327; 2019, pp. 27–28) where two friends, while sitting in a café, interrogate the cultural imaginations

they bestow on the people they see around them, the prejudices they represent, and the surface appearances that generate them. This then becomes an activity, not about noticing cultural differences, but about the origins of prejudice within the process of interculturality.

A further observation, in the first description, regarding the mistakes that AI can make, is that the three students are a mix of undergraduate and postgraduate which is unlikely if they are working on the same assignment. I also note that only the British student's city of origin is mentioned. This could be a form of Othering with regard to the Chinese and Egyptian students which is in line with the individualist-collectivist model in which only national or civilisational culture is relevant to their imagined identities. On the other hand, it could just be an innocent mistake where the reference to Manchester is just another part of the detail smoke-screen used to give the impression of authenticity.

The implication here is that ChatGPT could never be the 'neutral' interlocutor. On the other hand, no human interlocutor could be neutral. Because of the complexity of competing positions in any moment of small culture formation on the go, multiple and perhaps Othering and even self-Othering discourses will always be at play, as all of us are subject to the grand and personal narratives that surround us.

Conclusion

This has been an unusually difficult paper to write because every time I look at what ChatGPT produces it also takes me into an interrogation of the deep complexities of how it is possible to research and write about the intercultural—which I think also relates to how social science more generally is carried out and produced. I will therefore try to gain some clarification by summarising what I have discovered about AI with particular reference to the free version of ChatGPT, but also how this mirrors our own academic practice.

(1) A particular danger of using ChatGPT is that it can falsely appear to have direct engagement with the real world when in fact it cannot. At the same time, this reminds us of how we researchers also allow apparent and false engagement

with the real world which is falsely represented by naïve positivist methods and neoliberal structures. (2) The falseness of these methods of representation both with ChatGPT and our own practice can be hidden in detail. (3) ChatGPT will follow this established yet false line of representation until we challenge it, at which point it will more skilfully produce critical and deCentred arguments. We also need to work hard to challenge our own academic practice in order to achieve deCentred understandings. (4) However, no matter what ChatGPT appears to be able to do, it can never 'be there' in the settings it appears to represent. We therefore need to learn from this to do everything we can to demonstrate how we base our research on 'being there' in settings that are authentic to our experience. (5) Seeing that what ChatGPT produces is not 'real' should heighten our awareness of 'what is real' in our own writing.

It therefore becomes apparent that AI indeed 'mirrors our ideological blind spots' and, when we realise this, 'pushes us toward sharper criticality and reflexivity'; and accusing it of being biased 'without entering into critical and reflexive conversations with it' is 'unfair' (Dervin, 2026, pp. i, 11).

As an epilogue, I uploaded a nearly finished draft of this paper to ChatGPT and asked it to draft and abstract and a conclusion. What it instantly produced looked competent on the surface; but I could not use them because they did not quite say what I thought the paper was about. I then began to coach it to improve the drafts, reminding it to write in the style of which I approve, asking it to find examples of my successfully published abstracts. An objective of this activity was to help me find out if there were points I thought I had made in the paper which in fact I had not. This then helped me edit the paper to bring out more clearly the main points I wanted to make, and to write my own conclusion. Here ChatGPT is not being asked to write for me, but instead being asked as a research and writing assistant to help me write—with full knowledge of all its faults and also mine. After editing the whole paper and uploading again, I asked for another version of the abstract. It was not as good as I could do, I guess because there are other things out there which it was using which I would not; but at least a better starting point. What I will do when I have written my abstract, using its advice, but with my own knowledge of my academic community, I will upload that, keep the chat, thus educating my research and writing assistant to help better next

time. Eventually it might learn to write abstracts for me, but after huge input from my knowledge of the world I engage with to help it.

References

Amadasi, S., & Holliday, A. R. (2018). 'I already have a culture'. Negotiating competing grand and personal narratives in interview conversations with new study abroad arrivals. *Language & Intercultural Communication*, *18*(12), 241–256. DOI: 10.1080/14708477.2017.1357727.

Anderson, L. (2006). Analytic autoethnography. *Journal of contemporary ethnography*, *35*(4), 373–395. DOI: 10.1177/0891241605280449.

Atkinson, P. (2006). Rescuing autoethnography. *Journal of contemporary ethnography*, *35*(4), 400–404. DOI: 10.1177/0891241606286980.

Badwan, K., & Hall, E. (2020). Walking along in sticky places: post-humanist and affective insights from a reflective account of two young women in Manchester, UK. *Language & Intercultural Communication*, *20*(3), 225–239. DOI: 10.1080/14708477.2020.1715995.

Beny, R., & Nasr, S. H. (1975). *Persia: bridge of turquoise*. Thames & Hudson.

Canagarajah, A. S. (1993). American textbooks and Tamil students: Discerning ideological tensions in the ESL classroom. *Language, Culture and Curriculum*, *6*(2), 143–156. DOI: 10.1080/07908319309525145.

Canagarajah, A. S. (2018). Translingual practice as spatial repertoires: expanding the paradigm beyond structuralist orientations. *Applied Linguistics*, *39*(1), 31–54. DOI: 10.1093/applin/amx041.

Canagarajah, A. S. (2022). Challenges in decolonizing linguistics: the politics of enregisterment and the divergent uptakes of translingualism. *Educational Linguistics*, *1*(1), 25–55. DOI: 10.1515/eduling-2021-0005.

Clifford, J., & Marcus, G. E. (Eds.) (1986). *Writing culture: the poetica of politics of ethnography*. University of California Press.

Collins, H. (2018). Interculturality from above and below: navigating uneven discourses in a neoliberal university system. *Language & Intercultural Communication*, *18*(2), 167–183. DOI: 10.1080/14708477.2017.1354867.

Dai, D. W., Zhu, H., & Chen, G. (2025). How does interaction with LLM powered chatbots shape human understanding of culture? The need for Critical Interactional Competence (CritIC). *Annual Review of Applied Linguistics*, 45, 28–49. DOI: 10.1017/S0267190525000054.

Dervin, F. (2026). *AI for critical interculturality*. Routledge.

Fay, R., & Huang, Z. M. (2025). Intercultural musicking–portraits and narratives of klezmer ensemble performance. *Intercultural Communication Education*, *8*(2), 103051. DOI: 10.29140/ice.v8n2.103051.

Forster, E. M. (1909). The machine stops. *Oxford and Cambridge Review*, *8*(Michaelmas Term), 83–122.

Gasiorowska, K. (2026). *My reflections about the tortures and the pleasures of writing my PhD thesis*. Culture, Language & International Education Research (CLIER) Seminar, April 1st, Canterbury Christ Church University.

Geertz, C. (1993/1973). The interpretation of cultures. In: *The Interpretation of Cultures. Selected Essays*. Basic Books.

Hall, E. T. (1976). *Beyond culture*. Doubleday.

Hall, S. (1991). Old and new identities, old and new ethnicities. In: A. D. King (Ed.), *Culture, globalisation and the world-system* (pp. 40–68). Palgrave.

Hofstede, G. (2003). *Culture's consequences: comparing values, behaviours, institutions and organisations across cultures* (2nd ed.). Sage.

Holliday, A. R. (2012). Interrogating researcher participation in an interview study of intercultural contribution in the workplace. *Qualitative Inquiry*, *18*(6), 504–515.

Holliday, A. R. (2016). Difference and awareness in cultural travel: negotiating blocks and threads. *Language and Intercultural Communication*, *16*(3), 318–331. DOI: 10.1080/14708477.2016.1168046.

Holliday, A. R. (2019). *Understanding intercultural communication: negotiating a grammar of culture* (2nd ed.). Routledge.

Holliday, A. R. (2022a). *Contesting grand narratives of the intercultural*. Routledge.

Holliday, A. R. (2022b). Searching for a third-space methodology to contest essentialist large-culture blocks. *Language and Intercultural Communication*, *22*(3), 367–380. DOI: 10.1080/14708477.2022.2036180.

Holliday, A. R. (2025). Having a discussion with AI and the importance of direct experience. Retrieved from https://adrianholliday.com/having-a-discussion--with-ai-and-the-importance-of-direct-experience/. Access 9 June 2026.

Holliday, A. R. (2026a). ChatGPT as naïve but knowledgable research assistant. In: F. Dervin & H. R'boul (Eds.), *An introduction to AI and intercultural communication education* (pp. 11–27). Routledge.

Holliday, A. R. (2026b). What we all share about material culture. Retrieved from https://adrianholliday.com/what-we-all-share-about-material-culture/. Access 9 June 2026.

Holliday, A. R., & MacDonald, M. N. (2020). Researching the intercultural: intersubjectivity and the problem with postpositivism. *Applied Linguistics*, *41*(5), 621–639. DOI: 10.1093/applin/amz006.

Honarbin-Holliday, M. (2009). *Becoming visible in Iran: women in contemporary Iranian society*. I.B. Tauris.

Kebabi, A. (2024). 'Where are you from?' and 'foreigners': The discursive construction of identity in the personal everyday lives of well-established academics living in the UK. *Journal of Language and Discrimination*, *8*(1), 100–120. DOI: 10.1558/jld.25340.

Konik, R. (2025). Does art have to be created by humans? *Discourses on Culture*, *23*(1), 135–149. DOI: 10.2478/doc-2025-0007.

Kuhn, T. (1970). *The structure of scientific revolutions*. University of Chicago Press.

Lehman, I. M., & Canagarajah, A. S. (2024). Scholarly writing as a spiritual and reflexive practice. *Discourses on Culture*, *21*(1), 7–20. DOI: 10.2478/doc--2024-0001.

Liu, Q. (2026). *From soft dough to hard form: how the logic of 'Metal' (Jin) allows culture to travel*. Culture, Language & International Education Research (CLIER) Small Seminar: 'Doing qualitative research creatively: how to research a burning issue', February 20th, Canterbury Christ Church University.

Luckmann, T. (1978). Preface. In: T. Luckmann (Ed.), *Phenomenology and sociology*. Penguin.

Merabet, R. (2025). Students we label international: an urgent call to reconceptualise research with international students. *Equality, Diversity and Inclusion*, 44(3), 292–308. DOI: 10.1108/EDI-01-2024-0048.

Merrill, B., & West, L. (2009). *Using biographical methods in social research*. Sage.

Mills, C. W. (1970/1959). *The sociological imagination*. Pelican.

O'Regan, J. P., & Ferri, G. (2025). Artificial intelligence and depth ontology: implications for intercultural ethics. *Applied Linguistics Review*, *16*(2), 797–807. DOI: 10.1515/applirev-2024-0189.

Popper, K. (1945). *The open society and its enemies: Volume 1, Plato*. Routledge and Kegan Paul.

Ros i Solé, C. (2024). *Material Interculturality*. Routledge.

Said, E. (2003). *Orientalism*. Penguin.

Souleh, N. E. (2023). Creative nonfiction and stand-up comedy as alternative forms of decolonial scholarship. In: V. Odeniyi & G. Lazar (Eds.), *Decolonial Subversions*, *Special issue*, *Decolonising the university and the role of linguistic diversity* (pp. 31–47).

Talmy, S. (2011). The interview as collaborative achievement: interaction, identity, and ideology in a speech event. *Applied Linguistics*, *32*(1), 25–42. DOI: 10.1093/applin/amq027.

Triandis, H. C. (1995). *Individualism and collectivism*. Westview Press.

Weber, M. (1964/1922). *The theory of social and economic organisation*. The Free Press.

DOI: 10.2478/doc-2026-0006

This is an open access article licensed under the Creative Commons AttributionNonCommercial-NoDerivatives 4.0 International (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Jagat Bahadur Kunwar

School of Business and Economics, Åbo Akademi University, Turku, Finland

jagat.bahadurkunwar@abo.fi

ORCID ID: 0000-0002-2605-2760

Authenticity and Authorship Beyond the Algorithm: Action, Institution, and Epistemic Responsibility under Generative AI

Article history:

Received	**23 March 2026**
Revised	**6 May 2026**
Accepted	**7 May 2026**
Available online	**23 June 2026**

Abstract: This paper examines whether generative artificial intelligence can be considered an authentic author. While large language models (LLMs) demonstrate advanced linguistic competence and produce coherent academic discourse, they foreground unresolved questions concerning the ontology of authorship and the status of authenticity in scholarly authorship. The article is concerned with authorship rather than writing as such; writing is treated as the practical and institutional site through which authorship becomes visible, contested, and assigned.

Rather than treating authorship as a property of textual output, the paper conceptualises it as an activity across three intersecting perspectives: praxis, institutional recognition, and epistemic responsibility. Extending Hannah Arendt's The Human Condition, authorship is theorised in terms of linguistic labour, linguistic work, and linguistic action, with authenticity located in linguistic action as public, accountable, position-taking engagement. A Foucauldian perspective reframes authorship as an institutional function governed by discursive norms, while the epistemic perspective examines whether LLMs satisfy conditions such as semantic understanding, grounding, and epistemic commitment.

The paper argues that AI can perform linguistic labour, support linguistic work, and approximate institutional markers of authorship, but cannot initiate linguistic action or assume epistemic responsibility.

Keywords: authorship, authenticity, generative AI, large language models, epistemic responsibility

Introduction

Generative AI has made it increasingly difficult to treat authorship as a straightforward relation between a human subject and a written text. Large language models can produce fluent academic prose, reproduce disciplinary conventions, and generate texts that resemble scholarly argumentation. Yet these capacities do not by themselves settle whether such systems can be

authors. The central issue addressed in this article is not simply whether AI can write, but whether it can occupy the authorial position of attribution, responsibility, and answerability.

The classical benchmark for operationalising intelligence in an artificial system has been the Turing Test. In his "many minds" argument, Turing states that, since we cannot necessarily attribute intelligence through direct inward access even in other humans, it is unfair to deny it to machine systems on that basis. Hence, if the performance of a machine system is indistinguishable from that of a human through behavioural indicators such as those set up in the classical Turing Test, then the machine has performed indistinguishably from a human participant (Turing, 1950).

While the Turing Test was proposed in 1950, recent developments in artificial intelligence, especially generative AI such as various versions of ChatGPT, perform remarkably well on this criterion. For example, it has been shown that GPT-4 passes an interactive two-player Turing Test and was judged to be human 54% of the time (Jones & Bergen, 2024). Originally, Turing suggested that a system could be considered intelligent if it fooled human judges about 30% of the time in short conversations. More proprietary and presumably more advanced models have been introduced thereafter. Similarly, in evaluations of the Turing Test over 50 years, despite the inherent controversies surrounding it, it has been concluded that, while machines may not perfectly simulate human cognition, they possess sufficient cognitive capacities to pose ethical dilemmas in their use (French, 2000). For the purposes of this article, however, behavioural indistinguishability is not treated as sufficient for authorship. It merely sharpens the question of whether fluent, human-like textual performance can ground authorial responsibility.

This article is therefore concerned primarily with authorship rather than writing as such. Writing is treated here as the practical and institutional site through which authorship becomes visible, contestable, and assignable under conditions of generative AI. The article does not offer a general theory of academic writing or a comprehensive intervention into writing studies. Instead, it asks what distinguishes AI-assisted textual production from authentic authorship.

Recent debates on AI understanding, agency, and epistemic responsibility have moved beyond the Turing Test, acknowledging machine linguistic performance while asking whether symbolic manipulation in formal systems can amount to semantic understanding (Searle, 1980), whether computers can have subjective experience (Chalmers, 1995), how the architecture of LLMs relates to cognitive capability (Millière & Buckner, 2024; 2025), and what ethical problems follow from the use of generative AI in knowledge production (Hicks et al., 2024). These debates matter for authorship because authentic authorship, as understood here, requires more than behavioural indistinguishability or fluent textual output; it requires some relation to claims, reasons, evidence, commitments, and consequences.

Concurrently with these developments in debates on AI understanding, recent developments regarding authorship, still mostly confined to literary criticism, chart the historical evolution of the concept of authorship and conclude that authorship is primarily an affair of public rather than private consciousness, where the public domain consists of an "objective reality renderable by language, a shared vision of how the social should be restructured, or in terms of public conventions and traditions for the production and reception of discourse" (Burke, 1995, p. xviii). This observation also aligns with the "death of the author" argument proposed by Roland Barthes (1967), wherein the originality of the text is no longer attributed to a singular originating subject but understood as a blending and clashing of prior writings, themselves not original (Barthes, 1967). In this sense, the conventional figure of the author is displaced by the author as a "function of discourse", as articulated by Michel Foucault (1969), a shift that has subsequently shaped debates on authorship, particularly around questions of authorial intent, originality, autonomy, and aesthetic engagement (Puşcaşu, 2024). This is why the present article treats authorship not as mere textual production, but as a position constituted through public attribution, institutional recognition, and responsibility for claims.

These debates become newly consequential when LLMs can approximate many of the visible markers conventionally associated with authorship. At one level, the authorial enterprise is institutionalised through review processes,

templates, argumentative structures, citation practices, genre expectations, and disciplinary norms. These components are not simply matters of private consciousness but of the public and institutional domain. At another level, questions of authorial intent, originality, autonomy, and engagement bring us to questions of AI agency, emergent learning, understanding, and subjective experience. Generative AI therefore complicates not only writing, but also the institutional and epistemic conditions through which authorship is assigned, recognised, and made accountable.

Recent public instances of AI-generated work being submitted to or published in academic venues (e.g., Sakana.ai, 2026) sharpen the problem of authorial attribution. They suggest not simply that AI can generate academic-looking text, but that institutional markers of authorship may sometimes be approximated without the presence of a human authorial subject in the stronger sense of responsibility and answerability. More importantly, it may be the case that even human authors operate within institutionalised and disciplined modes of thinking that are, at least in part, imitable. As such, our creative enterprises may often be routine derivations of insights from given premises rather than sparks of human genius.

For example, Turing's reply to Lady Lovelace's objection, namely the objection that machines cannot originate anything and can only do what they are instructed to do, is instructive here. Turing asks "who could be certain that original work that he has done was not simply the growth of the seed planted in him by teaching, or the effect of well-known general principles" (Turing, 1950). For Turing, the Lovelace objection is weak because human originality itself may not be the wholly non-algorithmic enterprise we often assume it to be. Originality, rather than being fundamentally disruptive, may consist of a series of inductions combined with vast networks of information leading to surprising novelty. If this is so, then it raises the question: what kind of activity is authorship, what role do human authors play in this authorial enterprise, and how much can be safely attributed to LLMs?

These developments raise the central question of this article: can AI be regarded as an authentic author? The question is not whether AI can produce academic prose, but whether it can occupy the authorial position of responsibility,

attribution, and answerability. To address this question, the paper develops three complementary perspectives on authorship: praxis, institutional, and epistemic.

The first perspective draws on Hannah Arendt's distinction between labour, work, and action in The Human Condition (Arendt, 1958). Labour refers to repetitive, necessary, and consumptive activity; work refers to the production of relatively durable artefacts; and action refers to speech and deed through which persons appear before others, disclose who they are, and become answerable in a shared world. Translated into the present argument, generative AI can perform linguistic labour, such as drafting, summarising, paraphrasing, and stylistic smoothing. It can also support linguistic work, such as the production and organisation of durable scholarly artefacts. However, it cannot itself perform linguistic action, because it does not appear as an embodied and accountable subject before others.

The praxis perspective asks what kind of human activity authorship is and where, if anywhere, authenticity resides. The institutional perspective asks how authorship is authorised, attributed, and regulated within scholarly discourse. The epistemic perspective asks whether current LLMs can satisfy conditions associated with authorship, including semantic understanding, grounding, agential stability, and epistemic commitment.

The article addresses debates on AI authorship, scholarly attribution, research integrity, and epistemic responsibility in AI-mediated knowledge production. By integrating these three perspectives, the paper contributes to debates on AI authorship by clarifying why fluent textual production is not equivalent to authentic authorship. As will be apparent from the following sections, recent developments in AI collapse commonly understood proxies for voice, such as prose quality, forcing a re-examination of the authorial enterprise, authorial voice, and authenticity in scholarly discourse. Similarly, to the knowledge of the author(s), Arendt's ontology of human action has not been extended to authorial activity in this specific context. This paper therefore uses Arendt's labour–work–action distinction to differentiate linguistic labour, linguistic work, and linguistic action, thereby making it more practical to distinguish between AI and human roles in the authorial enterprise.

Additionally, the paper proposes a schematic framework for understanding authorship from three complementary but multidisciplinary perspectives: the praxis perspective, the institutionalised perspective, and the epistemic perspective. The framework positions authenticity in authorship at the nexus of public responsibility, institutional recognition, and epistemic standards. Building on these arguments, we distil a set of characteristic features that, together, clarify the debate on authenticity in authorship in relation to LLMs.

Furthermore, the paper develops the schematic analytically to distinguish authentic, performative, marginal, and simulated forms of authorship, thereby situating AI-generated and AI-assisted texts within a broader typology of authorial practice. More specifically, the contribution lies in integrating an Arendtian ontology of human activity with an institutional and epistemic account of authorship, and in developing a schematic that not only conceptualises these dimensions but also analytically differentiates forms of authorship and locates AI-assisted textual production within this structure.

The paper is structured as follows. First, it develops an Arendtian account of authorship as linguistic labour, linguistic work, and linguistic action. Second, it examines authorship as an institutional function through which scholarly authority, attribution, and accountability are organised. Third, it considers relevant debates on AI understanding, grounding, subjective experience, and epistemic commitment. The final section integrates these perspectives into a schematic account of authentic authorship under conditions of generative AI and briefly outlines directions for future research.

Authorship as praxis: linguistic labour, work, and action

Referring to Hannah Arendt's The Human Condition (Arendt, 1958), which distinguishes the ontology of human activity into labour, work, and action, we may conceptualise authorship as a human activity rather than merely the production of text. Arendt refers to labour as human activity comprising activities that are repetitive, necessary, and consumptive. Work, as opposed to

labour, is the production of durable artefacts, dealing with objectification and world-building activities. Lastly, action comprises activities that necessitate the disclosure of the "who", plurality, unpredictability, and the publicness of the enterprise. This section therefore treats writing only insofar as it materialises different forms of authorial activity.

If we recast the labour–work–action distinction in relation to authorship, these components become visible in the practices through which scholarly texts are produced. For example, much of the authorial enterprise, such as drafting boilerplate, summarising literature, reproducing genre norms, and iterative optimisation of the draft for clarity and fluency, including smoothing prose, paraphrasing, and summarising, can be considered "linguistic labour" because, as per Arendt's definition (Arendt, 1958; Saeidnia & Lang, 2017), these activities are repetitive, rule-governed, and oriented towards efficiency and academic consumption. It is important to note that AI performs linguistic labour extraordinarily well and is structurally strong precisely for these reasons. Linguistic labour, in this sense, does not by itself ground authentic authorship, because labour is not where authenticity resides in Arendt's account.

If we think about the conceptualisation of "work" in Arendt's sense, then we can say that "linguistic work" is about producing durable textual artefacts such as articles and reviews, stabilising arguments into objects, creating conceptual architecture, methodological inscriptions, and overall contributing to the archive of scholarship. In essence, linguistic work may be defined as scholarly artefact--making. AI also assists here, but some components of linguistic work still require human purposive restructuring and evaluative judgement, perhaps in the form of iterative prompts. It may be suggested that academic work is not always oriented towards action in the Arendtian sense, and the role of AI becomes problematic primarily when work is conflated with action. The point is therefore not to classify all writing as authentic or inauthentic, but to distinguish the forms of authorial activity that AI can assist from those it cannot occupy.

If we further use Arendt's notion of action (Arendt, 1958; Saeidnia & Lang, 2017) and define "linguistic action" as something that involves taking responsibility for a position, inserting oneself as an author into a contested public space, such as conferences and seminars, and risking disagreement,

reinterpretation, and response, then, in this sense, authorial voice is more than style. Linguistic action is a public, risky, identity-disclosing intervention into a plural space, taking responsibility for a position and initiating contestation. Voice is therefore about answerability and not merely fluency or prose quality. Linguistic action involves appearing as an embodied actor before others, presupposes plurality, and requires intelligibility to others. In this regard, AI cannot act in the true Arendtian sense because it cannot be answerable in the human public realm. Therefore, authorial voice is not merely a stylistic signature, but rather the residue of actions that involve accountability, risk, and positioning, even within linguistic work and labour.

Based on the labour–work–action distinction, and judging by the current capabilities of AI, we can claim that AI can perform linguistic labour remarkably well and assist greatly, or even fully, in linguistic work, but cannot initiate linguistic action. Only humans are able to do that by assuming public responsibility for meaning. We can conclude at this stage that when we talk about authorial authenticity or voice, it denotes responsibility for one's claims, errors, and explicit positioning, rather than merely a recognisable textual style. An author is responsible for the claims they make and the errors they commit, and is accountable for their explicit positioning, signifying that authentic authorship aligns with Arendtian linguistic action within the labour–work–action distinction. This conceptualisation explicitly resists a reductionist view of authorship as mere output and treats it as an activity.

However, such responsibility is not exercised in isolation but presupposes a shared epistemic space within which claims can be evaluated and contested. To add to the above observation, Chakrabarty's (1998) concern with the limits of epistemic pluralism highlights a condition often overlooked in debates about authenticity and voice. Albeit in the context of minority and subaltern histories, while they rightly contest dominant narratives, Chakrabarty warns that the rejection of shared criteria of fact and evidence risks undermining the very possibility of adjudicating competing claims in public life. This insight complements an Arendtian understanding of action as appearing before others in a shared world: without minimal agreement on what counts as evidence, speech ceases to be action and becomes either private expression or irreconcilable

assertion. Authenticity, therefore, cannot be equated with unrestricted self--expression; it presupposes participation in a common epistemic world that renders claims contestable and answerable. This also clarifies that truth is not treated here as privately possessed or politically neutral, but as publicly contestable within shared epistemic practices.

Authorship as an institutional function

In academic publishing, authorship is not merely a description of who produced a text. It is an institutional mechanism through which credibility, authority, ownership, responsibility, and eligibility to speak are distributed. Journals, peer review, citation practices, affiliation, editorial judgement, authorship guidelines, and disciplinary conventions help determine whose claims become recognisable as scholarly contributions and whose voices remain peripheral or unauthorised (Cronin, 2001; Chartier, 2003; Hyland, 2009; COPE Council, 2019). In this sense, much of the ontology of authorship is not merely psychological but procedural. Examples include the peer review process, citation practices, journal styles, and accountability mechanisms for authenticity and originality. Often, it is institutions that define what counts as meaningful, reproducible, and credible speech. The preoccupation of an author with analysing various journal metrics, such as citation density, genre conventions, and reviewer expectations, is already part of these authenticity-producing mechanisms. If we look closely, it becomes apparent that AI does not "kill authenticity" but exposes how little authenticity may be required for institutional recognition.

Similarly, as authors we are involved in the enterprise of reproducing or stabilising discourses, for example, the literature review in this paper, and, to some extent, challenging the discourse. If we consider LLMs, they are already capable of sedimenting past discourses and reproducing statistical crystallisations of academic norms, because they are trained on large corpora that include academic and academic-like texts. Effectively, the training data in LLMs already encode prior regimes of truth.

Following Foucault on discourse and the author-function, we may consider an author not simply as a person but as a regulatory node where sedimented discourse and crystallised academic norms precede and exceed the individual writer. The author-function is embedded within legal and institutional arrangements that delimit and organise the field of discourse. Its operation is historically and culturally variable, rather than stable or universal across contexts. It does not arise from a straightforward attribution of a text to an originating subject, but is constituted through a set of structured and often implicit procedures that govern how authorship is assigned. As such, it cannot be reduced to, or equated with, a concrete individual (Foucault, 1969; Chartier, 2003; Pușcașu, 2024).

If, then, authenticity is institutionally defined, and if AI-produced texts largely satisfy those institutional criteria, the objection to AI on the grounds of inauthenticity is partly jurisdictional rather than purely ontological: it concerns who is authorised to occupy the position of a speaking subject, who may occupy the author-function, and who can be held answerable. In other words, the question concerns authority and responsibility as much as textual production.

For the purposes of this article, the key point is that a Foucauldian understanding of authorship shifts authenticity away from personal sincerity and towards authorisation, recognition, and responsibility within institutional procedures. A systematic analysis of authorship as an institutional function therefore requires attention to the author-function, regimes of truth, and discursive authority, rather than to personal sincerity alone. The concern with AI is that it already incorporates many of the procedural logics through which authenticity is typically assessed, thereby blurring, and in some cases dissolving, the distinction between the authentic and the non-authentic at the level of institutional recognition.

This shift away from authenticity as interior sincerity towards its procedural and publicly recognisable dimensions also opens onto a complementary, non-interiorist account of authenticity. In this regard, Charles Guignon (2004) reorients the debate away from romantic self-expression and the metaphysics of psychological depth, and instead understands authenticity as situated,

practice-bound, and answerable to shared norms. From this perspective, authenticity does not require that a text be produced "from inside" a sovereign self, but can obtain without any strong commitment to interiorism. Rather, authenticity is practical, relational, and institutionally situated, not solely private or reducible to institutional approval, but concerns how an author inhabits norms without being wholly subsumed by them. Authentic scholarly voice is therefore not a distinctive style, but a pattern of responsible self-ownership in making claims, where the author can give reasons for their position, acknowledge the standards they are using, and stand behind the consequences of those commitments within a community of inquiry.

Bringing in the Foucauldian perspective on the author-function (Foucault, 1969; Chartier, 2003; Pușcașu, 2024) and Guignon's view of authenticity (Guignon, 2004) as practice-bound rather than a metaphysical property, we can conceptualise authorship not as an inner property of a subject, the author in this case, but as a discursively authorised classification. In other words, authenticity is recognised rather than expressed, voice is not possessed but validated, and originality is, by definition, institutionally ratified rather than grounded in private consciousness. Thus, the institutions that define authenticity may include editors, reviewers, disciplines, and other gatekeepers in the publication process.

This discussion suggests that authorship is partly an institutional function, extending to questions of attribution, accountability, credit, and disclosure. If authorship is an institutional function and is deeply involved in discourse reproduction, AI can act as an amplifier of genre norms and stabilise disciplinary style to a greater extent. However, once again, AI is limited in terms of epistemic responsibility when it comes to the accountability of the author regarding claims, errors, and positioning.

The point, closely related to the Arendtian analysis, is that debates about authenticity often misidentify linguistic labour or linguistic work as linguistic action. It is important to note here that we subscribe neither to total institutional determinism in the Foucauldian sense nor to the naïve humanism that might be suggested by Arendt alone in understanding authenticity. Both perspectives are required, as authorship is neither totally institutionally determined nor a heroic labour in which authentic voice emerges from expressive originality

alone. Whereas Foucault shows who defines authenticity in the sense of what makes a claim admissible as authentic knowledge, Arendt identifies what kind of linguistic activity is at stake, albeit without specifying epistemic criteria for intelligibility.

Epistemic perspectives on AI and capabilities

The final perspective concerns AI understanding, grounding, agency, and epistemic commitment. These debates matter for authorship because authentic authorship, as defined here, requires more than the production of fluent text. It requires some relation to meaning, reasons, evidence, claims, and consequences. The first issue concerns whether current LLMs "understand" language, a notion for which there is no settled or widely agreed meaning, especially given that authorship is primarily a linguistic enterprise. The second issue concerns whether LLMs can have "subjective experience", again a concept marked by competing interpretations and no definitive consensus, as our definition of linguistic action assumes accountability in public as a marker of authenticity, which in turn presupposes some form of agency and epistemic norms derived from authorial subjectivity.

On "understanding" language

The issue of understanding matters here only insofar as it bears on authorship: fluent linguistic production does not automatically entail authorial understanding, commitment, or responsibility. We must first clarify what is meant by "understanding language". In human cases, understanding presupposes an embodied speaker, intentions behind speech acts, and the ability to relate linguistic expressions to causal realities. When we say that LLMs do not understand language, we are making a claim that goes beyond mere cognitive competence: current LLM behaviour does not amount to the kind of understanding humans exhibit. The key distinction is between formal linguistic competence and pragmatic understanding. Producing well-formed

or convincing sentences is not sufficient; understanding involves grasping the meaning of those sentences in relation to the conditions from which they arise.

Our position aligns with the distinction between syntactic performance and semantic understanding. Searle's Chinese Room argument (Searle, 1980) shows how a system can generate appropriate linguistic outputs by manipulating symbols according to rules without grasping their meaning. The argument imagines a person manipulating Chinese symbols according to formal rules without understanding Chinese, thereby showing how appropriate linguistic output may be produced without semantic comprehension. From the outside, the system appears to understand the language, but internally it performs only formal operations. Applied to contemporary LLMs, fluent and contextually appropriate responses may therefore reflect syntactic pattern manipulation rather than genuine semantic comprehension. Related critiques, such as the blockhead argument and the stochastic parrot argument (Bender et al., 2021; Millière & Buckner, 2024), similarly suggest that systems trained to predict sequences of tokens can reproduce convincing linguistic outputs without possessing human-like understanding. If authorship is understood as accountable activity rather than textual output, this distinction is crucial.

Rejecting Turing's equivalence between intelligence and behavioural performance (Turing, 1950), Searle argues that genuine understanding requires intentionality, that is, mental states directed towards objects or states of affairs (Searle, 1980). LLMs manipulate symbols according to formal rules without attaching meaning to them. While such systems may generate appropriate outputs, the process remains purely syntactic. Understanding, by contrast, requires semantic relations to the world. The Chinese Room argument therefore suggests that rule-governed symbol manipulation alone cannot produce genuine understanding.

Understanding may also be linked to embodiment, where meaning emerges through sensorimotor interaction with the world. For authorship, embodiment matters because claims are made from situated positions and can be answered, challenged, or sanctioned within a shared world. Text-only systems lack such embodied cognition (Searle, 1980). This concern connects to the broader issue

of grounding. Understanding requires linguistic symbols to be connected to entities and causal relations in the world. Although LLMs show strong syntactic performance and may encode inferential relations between expressions, this does not guarantee referential meaning. Trained primarily on textual data rather than interaction with the world, any world models they possess emerge indirectly from linguistic patterns. As a result, LLMs may exhibit inferential semantic competence while lacking grounding, referential semantic competence, and robust representations of causal structure (Millière & Buckner, 2025). Describing model errors as "hallucinations" can therefore be misleading. Hallucination implies misperception, which risks anthropomorphising the system (Kang & Shaw, 2024), whereas LLMs do not perceive the world at all. Their outputs arise from probabilistic pattern generation rather than misperception, reflecting the same underlying limitations of grounding, reference, and causal connection to the world (Hicks et al., 2024). This is especially relevant where authorship involves situated forms of knowledge, since the author's relation to a world, community, or position can shape what is being claimed and what is at stake in making the claim.

Human language use is embedded in communicative intentions and broader patterns of action and social interaction. The meaning of linguistic expressions often depends on a speaker's intention within a communicative act, where language is used to achieve goals such as coordinating activity, expressing beliefs, or sharing information about the world (Millière & Buckner, 2025). By contrast, current LLMs are optimised for linguistic plausibility (Hicks et al., 2024), generating plausible continuations of text rather than participating in communicative action. Although their outputs may appear syntactically and semantically convincing, they are not grounded in genuine communicative intentions or practical engagement with the world, which raises doubts about whether they reflect genuine language understanding. If linguistic action is the hallmark of authentic authorship, as outlined previously, LLMs, devoid of communicative intentions, cannot necessarily participate in linguistic action.

Human communication presupposes agents with relatively stable beliefs, viewpoints, and commitments over time. This stability helps maintain consistent meanings in dialogue. By contrast, LLMs readily change positions when

prompted and can be induced to defend contradictory claims. This malleability suggests that their outputs are not anchored in a stable agent perspective, which undermines the idea that their utterances carry determinate meanings in the way human linguistic expressions do (Millière & Buckner, 2025). An authentic author probably has some degree of agential stability over time, which the current generation of LLMs lacks (Hicks et al., 2024).

This observation also connects to debates on the conditions under which something may be regarded as an autonomous agent. While some accounts attribute a form of instrumental agency to LLMs (Yildirim & Paul, 2023), others argue that such systems cannot be said to know anything precisely because they lack agency (Goddu et al., 2024). For the present argument, however, the key issue is narrower: whether LLMs can occupy the authorial position of responsibility, commitment, and answerability. Related debates on AI agency and free will remain unsettled, but what is at stake here is not LLMs as such, but how agency and knowledge are conceptualised in relation to authorship.

Furthermore, human language develops through cultural ratcheting, where speakers build on and stabilise the knowledge and practices of previous generations. This process involves not only producing linguistic outputs but also interpreting, evaluating, and integrating them within a shared cultural context. Although LLMs can generate sophisticated responses, they do not participate in this process of cultural transmission. Humans remain responsible for interpreting and stabilising their outputs, which suggests that LLMs do not meaningfully participate in cultural learning (Millière & Buckner, 2025). If we assume that, as authors, we participate in shared meaning-making within a given cultural context, LLMs could not be considered authentic authors.

Human linguistic assertions operate under norms of truth and epistemic commitment, where speakers present statements as representing how things are in the world. By contrast, LLMs generate sentences without any intrinsic orientation towards truth. This does not imply that human authors possess truth in an unmediated or politically neutral way. Scholarly truth is relationally produced through evidence, interpretation, citation, peer review, contestation, and correction. Precisely because truth claims are socially and institutionally mediated, authorship matters: claims require accountable subjects who can

be questioned, challenged, corrected, and held responsible. LLMs, by contrast, produce text that is statistically plausible given the preceding context. Consequently, when LLM outputs are correct, this accuracy often arises as a statistical by-product of patterns in training data rather than from an attempt to track or represent facts about the world. This truth-indifference distinguishes linguistic plausibility from genuine assertion. Although LLMs exhibit strong syntactic fluency and can track inferential relations between sentences, their outputs remain independent of the epistemic norms (Chakrabarty, 1998) that govern human linguistic practice. Linguistic competence in such systems therefore operates at the level of form and inference rather than referential meaning or epistemic commitment to truth. In this respect, their outputs may be characterised as "bullshit", in the sense of being produced without regard for truth (Frankfurt, 2005; Hicks et al., 2024). Although this is a normative commitment, as authors we cannot risk truth-indifference in academic enterprise, and so we may also add epistemic commitment as an ideal characteristic of authenticity.

Another limitation concerns the architecture of current language models. Most LLMs primarily model formal patterns in language rather than the broader cognitive systems that support human linguistic competence. Human language use is embedded in a wider set of psychological capacities, including perception, memory, attention, imagination, social cognition, and long-term planning, which help maintain coherent and systematic behaviour across contexts. Because current LLMs lack such an integrated cognitive architecture, their linguistic performance does not reflect the same kind of systematic understanding found in human language users (Millière & Buckner, 2025). Although LLMs may produce syntactically and semantically plausible texts, the systematic understanding and articulation of a given phenomenon, which is the subject of any academic article, remain limited and questionable.

Even if LLM performance approaches that of human speakers and appears to "understand" language, there are methodological concerns about such evaluations. In many contexts, their responses are indistinguishable from human outputs, which makes behavioural performance an appealing indicator of understanding (Jones & Bergen, 2024). However, evaluation benchmarks can become targets of optimisation, models may benefit from data contamination

because benchmark questions appear in training datasets, and there are deeper concerns about construct validity. Many benchmarks were designed to assess human cognition and rely on assumptions about how humans process language. As a result, strong behavioural competence on these tests does not necessarily establish that the system genuinely grasps meaning (Millière & Buckner, 2025). If this is the case, this point largely softens the debate on authenticity, as the crucial claim is that, without understanding language, one cannot produce authentic authorship.

A strong counter-argument to our position concerns compositionality in language (Millière & Buckner, 2025). Traditionally, linguistic competence has been understood as rule-based, grounded in explicit grammatical structures such as those proposed in theories of Universal Grammar by Noam Chomsky (Chomsky, 1986). However, the success of LLMs challenges this assumption. These systems produce grammatically well-formed and systematically related sentences despite lacking any explicitly encoded symbolic grammar. This suggests that productivity and systematicity in language may emerge statistically from large-scale pattern learning rather than from rule-based cognitive structures. If compositional structure can arise from patterns of linguistic usage alone, then the symbols and rules posited in classical linguistic theories may be descriptive tools rather than fundamental components of human cognition. On this view, strong linguistic competence might not require the kinds of cognitive mechanisms traditionally associated with language understanding, which weakens the claim that LLMs lack genuine linguistic competence. If emergent linguistic acquisition and understanding are possible through compositionality, a still largely debated issue, this can only be verified empirically, and so the position on language understanding may shift in the future, changing the dynamics of the concepts explored in this paper.

Thus, the point of this section is not to settle the philosophy of language or cognition, but to clarify why linguistic fluency alone cannot ground authentic authorship without understanding, grounding, and epistemic answerability.

Subjective experience, embodiment, and authorial position-taking

The relevance of subjective experience to this article is limited but specific. If authentic authorship involves position-taking, embodied exposure, and answerability, then subjective experience matters not as an abstract property of consciousness but as part of the broader question of whether an author can appear before others, be addressed, and bear consequences for claims. By "subjective experience", we refer to first-person phenomenal awareness, or the felt quality of experience (Nagel, 1974; Chalmers, 1995).

Our position is agnostic regarding machine consciousness. Current evidence does not justify attributing subjective experience to artificial systems such as LLMs, although some philosophical approaches, particularly computational functionalism, leave open the possibility that sufficiently advanced artificial systems could possess conscious experience in principle. The most defensible position is therefore neither to attribute nor categorically deny machine consciousness, but to ask whether current LLMs can satisfy the conditions relevant to authentic authorship.

Turing's discussion of the "argument from consciousness" remains relevant here. The objection claims that a machine cannot truly think because it lacks feelings or subjective experience. Turing responded that this requirement is problematic, since we cannot directly access the experiences of other minds, including those of other humans, and therefore typically infer mental states from behaviour (Turing, 1950). This reply cautions against a simple dismissal of machine consciousness, but it also shows the limits of behavioural evidence: convincing first-person statements generated by an LLM do not by themselves establish subjective experience.

As discussed above, Searle's Chinese Room argument supports the narrower point that formal symbol manipulation does not by itself establish understanding or intentionality, both of which are relevant to authorial responsibility. The discussion of grounding and world models also has implications for authorship. If subjective experience depends partly on embodied interaction with an environment, then purely text-trained systems lack at least some of the conditions associated with situated perspective. This does not settle the question of machine consciousness,

since the relationship between embodiment, cognition, and consciousness remains contested (Chalmers, 2023; Millière & Buckner, 2025). For the present argument, the point is narrower: authentic authorship presupposes an authorial position from which claims can be made, defended, revised, and answered for within a shared world.

Chalmers' (1995) distinction between the "easy" and "hard" problems of consciousness is useful here. The easy problems concern cognitive functions such as perception, information processing, verbal report, attention, and behavioural integration. The hard problem concerns subjective experience itself, or what Nagel (1974) describes as there being "something it is like" to be a conscious organism. This distinction matters because reproducing functional aspects of human communication does not by itself demonstrate that there is anything it is like to be an artificial system (Chalmers, 1995).

These debates remain unresolved. Functionalist accounts suggest that consciousness may depend on patterns of causal or informational organisation rather than biological substrate alone, leaving open the possibility of artificial consciousness in principle. By contrast, biological or embodiment-based accounts remain more sceptical about whether purely artificial or text-trained systems can instantiate subjective experience. For the purposes of this article, this debate need not be settled. The relevant point is that current LLMs do not provide sufficient grounds for attributing the subjective, embodied position--taking presupposed by authentic authorship.

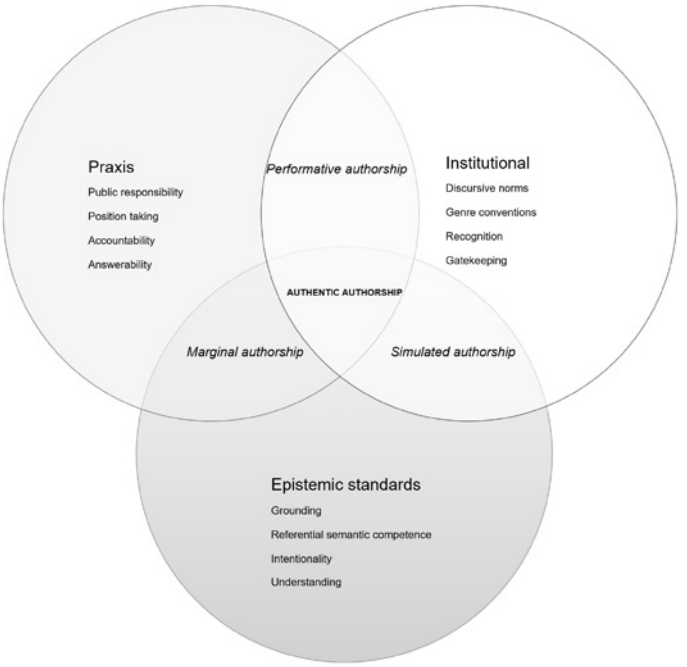
The above discussion illustrates that this remains a developing field and is relevant to authorship only insofar as subjective position-taking implies some form of embodied perspective, reflexivity, and answerability. At present, current evidence does not justify attributing such subjective experience to LLMs. This does not require a categorical denial of future artificial consciousness; it only means that current LLMs cannot be treated as authentic authors on this ground.

The schematic and discussions

By presenting three perspectives on authorship, namely the praxis, institutional, and epistemic perspectives, we can now move towards developing the idea of authenticity in authorship, particularly in relation to the existence and developing capabilities of LLMs. It is important to note here that the view of authenticity derived here is largely normative, as we claim it will always be, but it nonetheless contributes to debates on AI authorship, scholarly attribution, and epistemic responsibility.

Figure 1 shows the schematic representation of authorship as an intersection of praxis, institutional, and epistemic conditions, and illustrates how authenticity emerges at their convergence.

Figure 1. Authentic authorship as the intersection of praxis, institutional recognition, and epistemic responsibility



Source: Author's own elaboration.

While we identified the characteristics of authentic authorship as the nexus of praxis, institutional, and epistemic conditions, this does not indicate a moral hierarchisation of authorial ontology. These are also not three equal, independent, or modular domains, but rather interdependent conditions for authenticity in authorship.

The schematic also allows us to consider the analytical implications of partial intersections among these conditions. Different configurations generate further forms of authorship that clarify the position of AI-mediated textual production within the broader framework.

For example, when we combine the praxis and institutional dimensions, we observe a form of authorship that is recognised and publicly positioned but epistemically weak. An illustrative instance would be work that is institutionally validated yet displays indifference to truth. We label this as performative authorship.

Similarly, at the intersection of praxis and epistemic dimensions, we can infer a form of authorship that is grounded and involves genuine linguistic action, but remains institutionally unrecognised, often situated outside dominant discourse. We label this as marginal authorship, an instance of unrecognised authenticity.

Furthermore, if we combine the institutional and epistemic conditions, we observe a form of authorship that satisfies disciplinary standards of plausibility and institutional criteria, yet lacks the characteristics of praxis such as ownership, responsibility, and position-taking. This form may be described as simulated authorship, representing procedural epistemic production without linguistic action. It is important to clarify here that the epistemic dimension, as represented in the schematic, refers to the appearance of epistemic grounding rather than epistemic grounding in the stronger sense discussed earlier, such as intentionality, referential semantic competence, or genuine understanding of language. While AI systems can reproduce this appearance in terms of plausibility, coherence, and conformity to disciplinary standards, and appear "epistemic" enough to pass institutional filters, they do not satisfy the conditions required for grounding. In this sense, simulated authorship aligns with epistemic form without satisfying epistemic grounding, of which AI-mediated textual production is the most prominent contemporary instance.

In this way, the intersection of all three conditions stabilises authenticity. Rather than positing authorship as a simple dichotomy between human and AI, we conceptualise it as a set of analytical distinctions rather than a moral hierarchy, where different configurations are possible for human authors, while AI-mediated authorship most closely aligns with the intersection of institutional criteria and epistemic plausibility. In this sense, the schematic not only integrates existing theoretical perspectives, but also generates a typology of authorship that clarifies the position of AI within the broader ontology of scholarly practice.

Table 1 disaggregates these dimensions further by showing where current AI systems appear strong, where they provide partial support, and where they remain unable to satisfy the conditions of authentic authorship.

Table 1. Distribution of AI capabilities across the conditions of authentic authorship

Dimension	Authorial condition	AI capability	Authorship implication
Praxis	Linguistic labour	Strong	Does not ground authentic authorship
Praxis	Linguistic work	Partial/supportive	Requires human judgement and organisation
Praxis	Linguistic action	Absent	Requires embodied answerability
Institutional	Genre and procedural compliance	Strong	Can approximate authorial markers
Institutional	Attribution and accountability	Absent/derivative	Requires human or institutional responsibility
Epistemic	Plausibility and coherence	Strong	Can simulate epistemic form
Epistemic	Grounding and epistemic commitment	Weak/absent	Requires responsible authorial judgement

Source: Author's own elaboration.

As Table 1 indicates, AI systems can operate effectively at the level of linguistic labour and can support aspects of linguistic work, while remaining excluded from linguistic action. Similarly, AI can satisfy many forms of institutional procedural compliance, but it cannot assume attributional responsibility or be held accountable for claims. At the epistemic level, AI can reproduce plausibility and coherence, but it does not satisfy the stronger conditions of grounding and epistemic commitment discussed earlier.

The framework summarised in Figure 1 and Table 1 constitutes the central contribution of this paper. While existing debates on AI and authorship often proceed through abstract claims concerning creativity, intention, or intelligence, the present framework disaggregates authorship into analytically distinct yet interrelated dimensions and shows how AI intersects with each unevenly. In doing so, the discussion is shifted away from a binary question of whether AI can be considered an author, towards a more structured account of where and how authorship is distributed, simulated, or enacted within scholarly practice. The contribution of the paper, therefore, is not to resolve the question of AI authorship, but to render it analytically tractable by specifying the conditions under which authorship may be said to obtain, and by situating AI-assisted textual production within those conditions.

This analytical distinction should be read alongside the earlier Arendtian framing of authorship. Linguistic labour and many components of linguistic work can be performed or supported by LLMs, but linguistic action remains tied to public responsibility, position-taking, and answerability. Authentic authorship therefore resides most clearly in linguistic action: appearing before others, assuming responsibility for a claim, and becoming answerable within a shared scholarly world.

The institutional perspective adds that authorship is not reducible to an inner property of the writer. It is authorised through procedures, norms, and gatekeeping practices that determine who may speak, how claims become recognisable, and how credit and responsibility are assigned. AI can reproduce many of the procedural markers of scholarly authorship, including genre conventions and disciplinary style, but this does not by itself amount to authorial attribution in the stronger sense of accountability.

The epistemic perspective clarifies why fluent textual output is insufficient for authentic authorship. Current LLMs can produce plausible and coherent discourse, but they do not assume epistemic responsibility for the claims they generate. Whether future systems might satisfy stronger conditions of understanding, grounding, or subjective position-taking remains open; the present argument concerns the limits of current LLMs as authorial agents.

Authentic authorship does not require perfect fulfilment of all these conditions. Rather, it presupposes the capacity to take responsibility for claims, to relate those claims to a shared world, to sustain commitments over time, and to remain answerable within institutional and disciplinary practices. In this sense, authenticity is not a property of textual fluency but of responsible authorial positioning.

These positions sit comfortably with the view of an authentic author as neither merely sincere, expressive, political, nor moral, but as one who adopts a rationally defensible position within a shared epistemic framework (Chakrabarty, 1998). Plausibility is not understood here as access to absolute truth, but as intelligibility within historically and institutionally produced criteria of evidence, argument, and reality. Authenticity is therefore not unlimited self--expression; authorial action must remain epistemically legible and publicly contestable.

In summary, authenticity in authorship is situated at the intersection of Arendtian public responsibility, Foucauldian discursive authority, and epistemic defensibility. Whereas AI can assist linguistic labour and linguistic work, conform to institutional genre markers of authorial voice, and produce plausible academic discourse, it cannot be an authentic author in the strong sense developed here because it cannot own a stance, sustain commitments over time, or be accountable as an author embedded in scholarly practices. Authenticity in AI--mediated scholarly practice is therefore best understood as responsible self--ownership of epistemic commitments, enacted as public answerability within disciplinary criteria of plausibility.

The proposed framework also shows the interrelationships among these propositions. If authenticity is reduced to romantic self-expression, almost any linguistic output might count. If institutional authority alone is privileged, only

dominant voices count. If epistemic relativism is accepted, no adjudication is possible. For adjudication to remain possible, plurality and public responsibility must coexist with shared criteria of reality. This is why authentic authorship requires the intersection of praxis, institutional recognition, and epistemic responsibility.

AI does not threaten public reason simply by undermining facticity; it threatens it by displacing human responsibility for sustaining and contesting epistemic standards. It can stabilise the form of epistemic reasonableness, such as evidence, argument, citation, and coherence, without securing the substance of epistemic validity. AI hallucination illustrates the same distinction: a system may reproduce the form of scholarly plausibility while lacking grounding and responsibility for error. The concern, then, is epistemic automation without corresponding responsibility.

The conditions proposed here are not fixed once and for all. As AI capabilities develop, debates about understanding, grounding, agency, and subjective experience may shift. The question of authenticity in authorship should therefore be treated as dynamic rather than finally resolvable at a single technological moment.

Conclusion and future research

This paper has been primarily conceptual, with the aim of developing a framework for approaching the question of whether AI can be considered an authentic author. The argument has distinguished authorship from mere textual production by locating authenticity at the intersection of praxis, institutional recognition, and epistemic responsibility. From the Arendtian perspective, AI can perform linguistic labour and support aspects of linguistic work, but cannot itself initiate linguistic action. From the institutional perspective, AI can reproduce many of the procedural and genre-based markers through which authorship is recognised, but it cannot itself assume attributional responsibility. From the epistemic perspective, current LLMs can generate plausible and coherent discourse, but they do not satisfy the stronger conditions

of grounding, epistemic commitment, and public answerability required for authentic authorship.

Future research could examine how the distinction between linguistic labour, linguistic work, and linguistic action operates in concrete scholarly settings. One direction would be to analyse institutional AI policies as boundary-making texts that distinguish author from assistant, permissible support from inadmissible substitution, and procedural compliance from authorial responsibility. A second direction would be to examine drafting, peer review, and revision processes to identify where AI performs linguistic labour, where human judgement organises linguistic work, and where authors assume epistemic responsibility through linguistic action. Such work could use version histories, tracked changes, peer review reports, supervisor feedback, or author reflections to examine how responsibility and position-taking are enacted in practice.

Further work could also examine how disciplinary, institutional, and cultural contexts shape expectations of authentic authorship. The present article has not offered a comprehensive sociocultural account of authorship and authenticity; rather, it has developed a conceptual vocabulary for distinguishing textual production from authentic authorship as embodied, institutionally recognised, and epistemically answerable action. Since AI capabilities and institutional norms will continue to change, the question of authenticity in authorship should be understood as dynamic rather than finally resolved at a single technological moment.

References

Arendt, H. (1958). *The Human Condition*. Chicago: University of Chicago Press.

Barthes, R. (1967). The death of the author. In: *Image, music, text* (pp. 142–148). London: Fontana.

Bender, E. M., McMillan-Major, A., Gebru, T., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *FAccT*

'21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). DOI: 10.1145/3442188.3445922.

Burke, S. (1995). *Authorship: From Plato to the Postmodern (A reader)*. Edinburgh: Edingurgh University Press.

Chakrabarty, D. (1998). Minority histories, subaltern pasts. *Postcolonial Studies*, *1*(1), 15–29.

Chalmers, D. J. (1995). Facing Up To Problem of Consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219.

Chalmers, D. J. (2023). Could a Large Language Model be Conscious? *arXiv:2303.07103*. DOI: 10.48550/arXiv.2303.07103.

Chartier, R. (2003). Foucault's Chiasmus: Authorship between Science and Literature in the Seventeenth and Eighteenth Centuries. In: M. Biagioli, & P. Galison (Eds.), *Scientific Authorship: Credit and Intellectual Property in Science* (pp. 13–31). New York: Routledge.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.

COPE Council (2019). *COPE discussion document: Authorship*. Committee on Publication Ethics.

Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology*, *52*(7), 558–569. DOI: 10.1002/asi.1097.

Foucault, M. (1969). What is an author? In: *Language, counter-memory, practice. Selected Essays and Interviews* (pp. 113–138). Ithaca, New York: Cornell University Press.

Frankfurt, H. G. (2005). *On Bullshit*. Princeton: Princeton University Press.

French, R. M. (2000). The Turing Test: The First Fifty Years. *Trends in Cognitive Sciences*, 4(3), 115–121.

Goddu, M. K., Noë, A., & Thompson, E. (2024). LLMs don't know anything: reply to Yildirim and Paul. *Trends in Cognitive Sciences*, *28*(11), 963–964.

Guignon, C. (2004). *On Being Authentic: Thinking in Action*. Routledge: London and New York.

Hicks, M. T., Humphries, J., & Slate, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology*, *26*(38). DOI: 10.1007/s10676-024-09775-5.

Hyland, K. (2009). *Academic Discourse: English In A Global Context (Continuum Discourse, 7)*. London: Continuum International Publishing Group.

Jones, C. R., & Bergen, B. K. (2024). People cannot distinguish GPT-4 from a human in a Turing test. *arXiv:2405.08007*, 1–23. DOI: 10.48550/arXiv.2405.08007.

Kang, E., & Shaw, M. (2024). tl;dr: Chill, y'all: AI Will Not Devour SE. *arXiv:2409.00764*. DOI: 10.48550/arXiv.2409.00764.

Millière, R., & Buckner, C. (2024). A Philosophical Introduction to Language Models–Part I: Continuity With Classic Debates. *arXiv:2401.03910*. DOI: 10.48550/arXiv.2401.03910.

Millière, R., & Buckner, C. (2025). A Philosophical Introduction to Language Models–Part II: The Way Forward. *arXiv:2405.03207*. DOI: 10.48550/arXiv.2405.03207.

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, *83*(4), 435–450. DOI: 10.2307/2183914.

Puşcaşu, I.-d. (2024). Creating with AI: On recent debates about authorship revisiting the influence of Barthes and Foucault. *Studia Ubb. Philosophia*, *69*, 49–60. DOI: 10.24193/subbphil.2024.sp.iss.03.

Saeidnia, S. A., & Lang, A. (2017). *An Analysis of Hannah Arendt's The Human Condition*. London: Macat International Ltd.

Sakana.ai (2026, 28 January). Retrieved from https://sakana.ai/ai-scientist/.

Searle, J. R. (1980). Minds, brains, and programs. *The Behavioural and Brain Sciences*, *3*(3), 417–457. DOI: 10.1017/S0140525 × 00005756.

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *LIX*(236), 433–460. DOI: 10.1093/mind/LIX.236.433.

Yildirim, I., & Paul, L. (2023). From task structures to world models: What do LLMs know? *arXiv:2310.04276v1*. DOI: 10.48550/arXiv.2310.04276.

DOI: 10.2478/doc-2026-0007

This is an open access article licensed under the Creative Commons AttributionNonCommercial-NoDerivatives 4.0 International (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Ken Hyland

Education and Lifelong Learning, University of East Anglia, Norwich, United Kingdom

k.hyland@uea.ac.uk

ORCID ID: 0000-0002-4727-8355

Rethinking AI in Writing Pedagogy: From Fluency to Rhetorical Competence

Article history:

Received	**13 March 2026**
Revised	**25 March 2026**
Accepted	**25 March 2026**
Available online	**23 June 2026**

Abstract: Generative AI tools such as ChatGPT are reshaping academic writing in higher education, offering students unprecedented fluency, coherence, and accessibility. Yet these seductive characteristics, together with serious rhetorical

limitations, raise concerns about authorship, integrity, and the development of critical thinking. The similarities and differences between AI generated texts and human writing, and the implications of these for writing instruction, however, remain underexplored. This study aims to bridge this gap by synthesizing five studies conducted by Jiang and Hyland, which systematically compare ChatGPT--generated essays with those authored by British undergraduates. The studies reveal a consistent pattern: while ChatGPT excels in producing structurally polished texts, it underperforms in rhetorical depth, interpersonal engagement, and authorial presence. Building on these findings, the paper proposes a framework for integrating AI into writing pedagogy. This positions ChatGPT as a supportive tutor rather than a proxy author, stressing both its advantages and limitations. It offers practical pathways for educators and underscores the importance of cultivating critical AI literacy, enabling students to navigate new writing contexts where human creativity and algorithmic fluency intersect.

Keywords: ChatGPT, argumentative essays, AI literacy, written interaction, writing pedagogy

Highlights

- Five studies comparing ChatGPT and student essays reveal that AI excels in fluency and coherence but underperforms in stance, engagement, and rhetorical depth.
- I position ChatGPT as a scaffold, highlighting its strengths while recognising its limitations.
- I propose tasks such as contrastive analysis, revision of AI drafts, and prompt design to foster rhetorical awareness and critical AI literacy.
- I call for recalibrated rubrics and process-oriented practices to safeguard construct validity to assess student learning accurately.
- I emphasise transparency in AI use and the need to guide students to reflect on authorship, agency, and integrity.

Introduction

The rapid emergence of generative artificial intelligence (AI) tools such as ChatGPT has provoked widespread debate in higher education. On the one hand, these systems can produce fluent, coherent, and lexically sophisticated texts at remarkable speed, offering students accessible models of organisation and style. On the other hand, their rhetorical limitations—weak stance-taking, limited engagement, and formulaic cohesion—raise concerns about plagiarism, diminished critical thinking, and the erosion of academic literacy. The Guardian newspaper, for example, recently reported a 'freefall in public confidence' in US higher education with the potential for AI to increasingly substitute independent thought. Instructors at US Ivy League universities in the article believed that reliance on artificial intelligence was 'fundamentally antithetical to the development of human intelligence' (Speri, 2026).

As universities grapple with questions of integrity, pedagogy, and assessment, however, much of the discussion of AI in writing seems overly instrumental, preoccupied with detection, regulation, or efficiency (Gao et al., 2025). What risks being overlooked is the extent to which AI forces us to confront assumptions about the relationship of writing to learning and disciplinary participation. The main challenge, it seems, lies not in resisting AI outright but in understanding its capacities and limitations, and in integrating it responsibly into classrooms and other academic contexts.

This paper therefore examines AI in academic writing, focusing on student contexts and exploring how the distinctive features of AI generated texts can inform how we use these tools to develop literacy skills. It draws on five corpus-based studies conducted by Jiang and Hyland in the last 3 years, which systematically compare ChatGPT-generated essays with those written by British undergraduates. These studies analyse key rhetorical dimensions—stance and voice, cohesion and coherence, engagement and interaction, argumentation and evidence, and genre awareness to provide a comprehensive empirical base for evaluating AI's rhetorical performance. Together, they reveal a consistent pattern: while ChatGPT excels in producing structurally clear and fluent prose,

it lacks rhetorical penetration and interpersonal engagement and so fails to produce effective academic writing.

Building on these findings, I outline a conceptual framework for integrating AI into academic writing pedagogy. Rather than treating ChatGPT as a automated writer or a threat to integrity, the framework positions it as a scaffold for teaching rhetorical awareness and fostering critical AI literacy. By juxtaposing AI outputs with student texts, revising drafts to add rhetorical features, and embedding process-oriented assessment practices, educators can harness AI's strengths while tackling its limitations. The framework thus offers a roadmap for using AI creatively and critically in higher education, ensuring that students develop the metalinguistic and rhetorical skills necessary to navigate hybrid writing environments where human creativity and algorithmic fluency intersect.

Friend or foe? AI in Higher Education

Since its release in late 2022, ChatGPT has rapidly become one of the most debated tools in higher education, particularly in the domain of academic writing. Scholars have noted its ability to generate grammatically accurate, coherent, and lexically sophisticated texts, raising both opportunities and challenges for pedagogy and assessment. For many learners, particularly those writing in a second language, AI offers accessible models to assist leaners in organising their texts and polishing their grammar, functioning as a supportive assistant for developing academic literacy (Barrot, 2023; Deep & Chen, 2025). Studies also suggest that AI can support productivity by streamlining drafting, suggesting revisions, and assisting with citation formatting, thereby reducing the mechanical burdens of writing (Imran & Almusharraf, 2023). In this sense, then, ChatGPT acts as a helpful prompt or "writing assistant" capable of modelling academic genres and helping students visualise the rhetorical structures they might otherwise struggle to produce.

Yet alongside these opportunities, scholars have raised significant concerns. There is a risk, for example, that AI-assistance remains superficial, enabling students to meet formal requirements without gaining deeper control over

disciplinary meaning-making (Li & Wu, 2025; Siddiqui et al., 2025). One recurring theme is the apparent rhetorical shallowness of AI-generated texts, so while ChatGPT produces structurally polished essays, it consistently underperforms in authorial presence and interpersonal dimensions. This absence of rhetorical complexity not only produces emotionally bland texts but encourages students to rely on surface fluency rather than developing the critical thinking and persuasive skills that we value in academic writing. Ethical issues compound these concerns: questions of plagiarism, transparency, and authorship remain unresolved, with institutions struggling to balance innovation with integrity (Deep & Chen, 2025). These issues are compounded, and even stimulated, by assessment practices which use rubrics that privilege clarity and coherence and which, as a result, inadvertently reward AI outputs over student learning. This has prompted calls to recalibrate descriptors to emphasise reasoning, adaptability, and responsiveness to audience (e.g. Hyland, 2026).

Unfortunately, much of the discussion surrounding AI reduces a complex pedagogical issue to a question of rule compliance. From the perspective of L2 writing, the more consequential issue concerns authorship and responsibility (Li & Wu, 2025). Writing is not simply the production of text that meet external criteria; it is an act through which writers take responsibility for meanings, positions, and claims and which can engage and persuade readers. L2 writing specialists, then, are beginning to recognise that the challenge is to harness AI in ways that genuinely expand participation rather than merely smooth over inequality at the level of text. What this means is that whether AI works to narrows or widen existing proficiency and literacy gaps will depend less on the technology itself than on how we make use of it: on the pedagogical and institutional contexts in which it is embedded.

Recent scholarship has begun to explore how these challenges might be addressed pedagogically. Pack and Maloney (2023), for example, argue for teaching "prompt literacy", helping students understand how input shapes output and encouraging them to reflect critically on AI's rhetorical limitations. Similarly, Imran and Almusharraf's (2023) systematic review concludes that while ChatGPT can enhance academic writing, sustainable integration requires embedding critical AI literacy into the curriculum. This involves

not only recognising what AI can do well but also identifying what it cannot: constructing persuasive arguments, engaging readers, and signalling authorial stance. By positioning AI outputs as objects of analysis and revision, educators may be able to help students develop metalinguistic awareness and rhetorical competence.

In sum, the current literature exhibits an overall ambivalence towards AI in higher education. On one hand, ChatGPT offers powerful tools for modelling fluency and supporting students in the early stages of writing, while on the other, its rhetorical limitations and ethical risks demand careful pedagogical responses. This tension provided the rationale for the studies conducted by Kevin Jiang and myself examining AI's role in academic writing, drawing on parallel corpora to determine where its strengths and weaknesses lay. The results of these studies can help develop frameworks that integrate AI creatively and critically into university classrooms.

The need for comparison studies

The increasing sophistication of AI-generated texts raises pressing issues about academic integrity and potential effects on critical thinking skills. If students can easily access AI-generated essays that are indistinguishable from human--authored work, it not only undermines academic ethics (Barrot, 2023) but might impede the development of students' analytical and argumentative skills, which are crucial for their academic and professional growth (Adeshola & Adepoju, 2023; Kim et al, 2025). The issue is becoming critical as increasing numbers of students turn to AI to for assistance with their writing. A global survey by the Digital Education Council (Kelly, 2024), for example, found that 86% of students use AI in their studies, with 54% using it weekly, a rate of adoption mirrored in the Chegg survey of nearly 12000 undergraduates across 15 countries (Chegg, 2025).

Recent studies have therefore sought to identify features that can distinguish between AI-generated and human-authored texts. For example, AI texts seem to exhibit higher informational density with more noun and preposition-based structures (Herbold et al., 2023; Markey et al., 2024), while human-authored

texts exhibit a higher frequency of clause-related bundles and more pronounced use of epistemic stance markers (Jiang & Hyland, 2024). Yang et al. (2024) also found that ChatGPT-generated texts rely more on repetitive connective signals and constant thematic progression, suggesting less careful planning for text development, which leads to redundancy and a simplistic list-like structure. This can be attributed to the fact that the model leverages statistical analyses, large data collections, and machine learning to determine the likelihood of text sequencing rather than an understanding of language (e.g. Wolfram, 2023).

Texts created by ChatGPT's statistical algorithms therefore appear to reinforce a view of writing where it might be seen as a fixed structure following a system of grammatical rules (Vee, 2023). It tends to objectivize academic writing and so "fails to take account of the beliefs and knowledge writers assume readers will draw on in reading their texts" (Hyland, 2026, p. 6). The critical analyses of ChatGPT generated texts can therefore be helpful in providing linguistic insights into the limitations of the LLM in terms of how it understands textual interaction. This is particularly the case with argumentative writing, where writers seek to establish a coherent disciplinary context by drawing on their background knowledge and logical reasoning, and attempt to persuade readers to accept their position. Success is therefore a rhetorical outcome of what authors include in the text and the stance they take towards it (Hyland & Jiang, 2016; Aull & Lancaster, 2014).

Academic literacy is not simply a matter of producing grammatically accurate or structurally coherent texts; it demands an awareness of appropriate genre conventions and how these might be manipulated as well as the understanding and use of interactive and interpersonal aspects of argument that involve writers taking a stance and seeking to imagine and persuasively engage an audience. These are precisely the areas where generative AI tools such as ChatGPT remain underexplored. The corpusbased studies conducted by Jiang and Hyland provide a comprehensive foundation for addressing this gap, offering detailed analyses of rhetorical features across student and AI essays[1].

[1] These studies are listed in the references as Jiang and Hyland 2024, 2025a, 2025b, 2026, and in press.

Each of these studies examines a distinct rhetorical dimension of writing, including stance and voice, cohesion and coherence, engagement and interaction, argumentation and evidence, and genre awareness. Together, they offer a systematic comparison of ChatGPT-generated essays and student-authored texts across the core dimensions of academic literacy. The findings reveal consistent patterns: while ChatGPT excels in fluency and coherence, it underperforms in rhetorical depth, interpersonal engagement, and authorial presence. This evidence base is critical for moving beyond anecdotal debates about AI in education toward a grounded understanding of its rhetorical capacities. The rationale for focusing on these five studies is therefore twofold. First, they provide robust empirical evidence of how AI-generated texts differ from student writing in ways that are pedagogically significant. Second, they collectively point toward the need for a framework that integrates AI into academic writing instruction to help model rhetorical awareness and foster critical AI literacy. By situating AI within this framework, the present paper contributes to ongoing debates about the future of academic writing, offering a pathway for educators to harness AI's advantages while addressing its limitations.

The data and analyses

All five studies are based on the same two balanced corpora of argumentative essays designed to enable systematic comparison between student and AI-generated texts. Table 1 shows the characteristics of the corpora in terms of texts, words (tokens), and different words (types).

Table 1. Corpus characteristics

Corpus	No. of texts	Tokens	Types
British student essays	145	78,060	6,975
ChatGPT essays	145	72,819	7,164

Source: Author's own elaboration.

The first corpus comprised 145 essays written by British university students aged 18–22, drawn from the Louvain Corpus of Native English Essays (LOCNESS) (Granger, 2013). These texts were produced in untimed, coursework-like settings, reflecting a developmental stage in academic writing where rhetorical awareness and authorial identity are emerging. The topics—Britain's relationship to the EU, transport, boxing, and the lottery—are typical prompts in English for General Academic Purposes (EGAP) courses, encouraging students to take a stance and provide supporting arguments while also reflecting specific cultural and institutional perspectives on academic argumentation.

The second corpus mirrors the student one with 145 essays generated by ChatGPT in response to prompts following the student tasks as closely as possible. Each prompt specified a role ("competent university-student writer"), a realistic academic context, and a persuasive writing style. Prompts were refined to reduce ambiguity but deliberately avoided extensive optimisation, ensuring that outputs reflected ChatGPT's "default" rhetorical style under classroom--like conditions rather than artificially enhanced performance. For example, the transport prompt instructed:

> "You are a competent university-student writer of English texts for academic purposes. Write 16 argumentative essays with a persuasive writing style on the topic of [transport]. Each essay is about 500 words long".

Each AI essay was generated independently to maximise lexical and syntactic diversity and avoid any replication effects. This approach captures how learners might realistically engage with generative AI tools using broad thematic prompts.

The analyses mainly draw on Hyland's (2005) frameworks of stance and engagement and metadiscourse, which see academic writing as a form of interpersonal negotiation where writers project credibility, evaluative judgments, and alignment with readers. This framework has provided the basis for numerous studies across a range of genres, such as academic blogs (Zou & Hyland, 2020) and doctoral confirmation reports (Jiang & Ma, 2018) but also include argumentative essays (Aull & Lancaster, 2014). The study of metadiscourse has gained particular traction in comparing different groups of

writers, which reveal the challenges of meeting academic norms (Hyland, 2005; Hyland & Jiang, 2018). These have revealed contrasts in high- and low-scoring essays (e.g., Ho & Li, 2018), native and non-native speaker essays and genres (Hyland et al., 2022), but none have examined rhetorical performance in human and machine generated texts.

In this study all texts were standardised to a base of 1,000 words to enable direct comparison. Corpus linguistics tools were used to identify and quantify relevant features, with statistical significance tested using Rayson's (2016) log--likelihood calculator at the threshold of LL ≥ 3.8 ($p < .05$). To complement significance testing, effect sizes were calculated using %DIFF (Gabrielatos, 2018), which measures the magnitude of difference between normalised frequencies. This dual approach ensured that findings reflected both statistical robustness and practical relevance. Qualitative analysis supplemented quantitative results, illustrating how rhetorical choices shaped argumentation, authorial presence, and reader engagement.

Like all corpus-based research, there are certain limitations to these studies. The reliance on LOCNESS restricts the student sample to British undergraduates, which may not capture the diversity of academic writing across cultural and disciplinary contexts. Similarly, the ChatGPT texts reflect outputs from a specific version of the model at a particular point in time, meaning that findings may shift as AI systems evolve. The focus on short argumentative essays also limits generalisability to other genres, such as research reports or reflective assignments. Finally, while corpus methods reveal patterns of language use, they cannot fully account for the cognitive or social processes underlying writing, nor the pedagogical contexts in which AI is deployed. Despite these limitations, the methodological design provides a robust basis for comparison, offering empirical evidence of the rhetorical strengths and weaknesses of AI-generated texts.

Rhetorical differences: what do the results show?

The five corpus-based studies conducted by Jiang and Hyland provide a detailed picture of how ChatGPT-generated essays compare with those written by

undergraduate students. Taken together, they reveal a strikingly consistent configuration: while AI excels in producing fluent, coherent, and structurally polished texts, it falls short in the rhetorical dimensions that are central to academic literacy.

The first set of findings concerns ***stance and voice***. Student essays frequently employ hedges, boosters, and attitude markers to signal commitment, evaluation, and epistemic positioning. These features help construct a persuasive voice and establish authorial presence. By contrast, ChatGPT essays contain far fewer stance markers, resulting in a more impersonal tone and weaker rhetorical identity. This absence of evaluative language underscores the model's tendency to produce text that is grammatically correct but rhetorically flat (Hyland & Jiang, 2023). So, while LLMs offer considerable support for surface-level linguistic production, they do not yet simulate the complex rhetorical work that stance performs in mediating epistemic authority and academic identity. For educators, this opens up possibilities for pedagogical intervention: by positioning ChatGPT as a critical object of analysis, teachers can foster learners' understanding of how language choices shape argumentation, credibility, and engagement in academic discourse.

In terms of ***cohesion and coherence***, Jiang and Hyland explored regular, high frequency three-word sequences or 'lexical bundles'. Here, ChatGPT demonstrated strong structural clarity, often relying on additive transitions such as "moreover" and "in addition", as well as explicit stage-labelling phrases like "this essay will...". These devices ensure logical sequencing but are generally more formulaic than human writing and include very few devices which reformulate material for readers. The AI generated texts also exhibited greater lexical variation than the student essays although contained a lower frequency of bundles overall, suggesting that its bundles are more rigid and standardised. Clearly, ChatGPT is proficient at recognizing and generating fixed structures due to their high frequency and regularity in the language data it was trained with. Student essays, in contrast, while less consistently fluent, employed a wider range of cohesive strategies, including reformulation and topic-shifting markers, which contribute to rhetorical flexibility and nuanced argumentation.

The studies also highlight differences in ***engagement and interaction***. Student essays frequently use questions, personal asides, and direct appeals

to readers, creating a dialogic relationship and anticipating counter-arguments. ChatGPT essays, on the other hand, rarely employ such engagement markers, resulting in reduced interpersonal connection. A striking example of this difference is the near absence of personal asides in ChatGPT essays, where writers break off to make a personal comment on the ongoing text. This finding serves as a clear counterpoint to any tendency to anthropomorphize or "personify" the model. Despite generating coherent text, ChatGPT rarely inserted personal comments, questions or digressions as the human writers often did, underscoring its nature as a statistical model rather than a sentient entity. Although it was able to generate a series of extended and coherent responses to our prompts, it lacked the ability to adopt a strong perspective on a topic and to engage in persuasive interactions to carry it through, thus neglecting aspects of argument that are highly valued in academic writing. These limitations reinforce the model's tendency to produce monologic texts that lack responsiveness to audience.

Analysis of ***argumentation and evidence*** further illustrates the gap. The Lexical bundle analysis shows that ChatGPT essays are dominated by noun + preposition structures (e.g., "the role of", "the advantage of"), which provide abstract descriptions but little argumentative force. Student essays, by contrast, employ more clause-based bundles (e.g., "it is important to note"), which signal reasoning, evaluation, and authorial presence. This distinction once more highlights something of the rigidity of AI argumentation compared to the persuasive coherence of student writing. Again, this almost certainly is a result of the machine learning mechanism that gives ChatGPT its power. The massive dataset of texts which provides its learning input almost certainly contains a high density of noun and preposition-related bundles as these are common in many registers of writing.

Finally, the studies reveal differences in ***genre awareness***. As I noted earlier, ChatGPT consistently marks stages and transitions, producing structurally clear texts. Yet it lacks the interpersonal nuance provided by hedges, boosters, and attitude markers, which balance objectivity with authorial presence. Student essays, though less polished, displayed a wider repertoire of organisational

strategies, including consequential linking, topic shifts, and reformulation, which are indicative of purposeful reasoning and sensitivity to audience needs.

They also showed greater genre awareness by combining organisation with interpersonal positioning, thereby demonstrating a richer understanding of academic discourse.

These findings are summarised in Table 2.

Table 2. Comparison of AIGenerated vs. Student Essays

Dimension	ChatGPT essays	Student essays
Stance and Voice	Fewer hedges, boosters, and attitude markers Weaker authorial presence More impersonal tone	More stance markers Stronger authorial identity and persuasive voice
Cohesion and Coherence	Heavy reliance on additive transitions ("moreover", "in addition") and stage--labelling ("this essay will...") Formulaic cohesion	Wider range of cohesive devices, including reformulation and topic shifts Greater rhetorical flexibility
Engagement and Interaction	Limited use of questions, personal asides, and direct appeals to readers Limited reader connection	Frequent engagement strategies More dialogic interaction and anticipation of objections
Argumentation and Evidence	Dominated by noun + preposition bundles ("the role of") Abstract, rigid argumentation	More clause-based bundles ("it is important to note") Greater coherence using reasoning, evaluation, and persuasive devices
Genre Awareness	Strong structural clarity, Consistent stage marking Reduced interpersonal nuance	Combine organisation with hedges, boosters, and attitude markers Richer rhetorical awareness

Source: Author's own elaboration.

Across all five dimensions, the findings converge on a clear conclusion: ChatGPT produces fluent and coherent texts but lacks the rhetorical depth,

interpersonal engagement, genre awareness and authorial presence that define effective academic writing. Our findings indicate that AI-generated texts seem to have little difficulty in satisfying surface expectations of coherence, but fail to demonstrate the deeper competencies that written assessments are designed to measure. This means that if our assessment criteria overemphasise structural clarity or linear cohesion, they risk privileging machine-generated writing over the developmental evolution of student learning. Understanding these distinctions is therefore essential for preserving construct validity in writing assessment and for supporting principled responses to the increasing presence of AI in academic contexts.

While algorithmic coherence by LLMs facilitates logical flow, it cannot currently replicate the reader-driven process of constructing meaning that is characteristic of human-written texts. This limited rhetorical flexibility highlights the importance of interaction in crafting persuasive and engaging arguments. This distinction between surface fluency and rhetorical competence provides the empirical foundation for the pedagogic framework proposed in the next section. This framework employs AI as a scaffold for teaching rhetorical awareness and fostering critical AI literacy, rather than as a automated author. Overall, then, the results suggest that AI literacy should become an integral component of writing instruction.

Implications: Towards an AI-informed pedagogical framework

Essentially, the framework positions ChatGPT as a writing assistant. Its ability to deliver texts with grammatical accuracy, structural clarity, and lexical sophistication make it a valuable model for students struggling with fluency and organisation. However, its obvious limitations in interactivity and rhetorical nuance mean that it cannot replace the intellectual and interpersonal work of academic writing. Instead, AI outputs might better be treated as pedagogical objects of analysis, allowing students to identify rhetorical gaps and reflect on how language choices shape persuasion and credibility. The framework therefore

carries, I think, significant implications for pedagogy, assessment, ethics, and future research in writing instruction in higher education.

Pedagogically, it suggests that ChatGPT should be integrated into writing instruction as a scaffold for developing rhetorical awareness. Rather than treating AI outputs as finished products, instructors can frame them as texts to be interrogated, critiqued, and revised. In this way, students are encouraged to recognize the rhetorical gaps that distinguish fluent prose from persuasive academic arguments. Classroom activities might include contrastive analysis, where learners compare AI-generated essays with their own or with published exemplars to foreground differences in, say, stance and engagement. Teachers may use the model's outputs, as we have in these studies, as a contrasting backdrop to human writing, focusing on salient differences to foster skills of critical analysis, enabling students to discern between the mechanical organisation of ideas and the nuanced, persuasive expression that characterises good argumentative writing. Students can also be encouraged to improve upon these AI-generated drafts by adding personalized and contextually appropriate rhetorical features, thereby developing their skills in crafting more effective academic arguments.

In addition, revision tasks, where students enhance AI drafts by inserting stance markers, hedging devices, and engagement features, can also offer valuable opportunities for learners to cultivate authorial presence and metalinguistic awareness. Prompt design experimentation by students further encourages exploration in how input shapes output, fostering metalinguistic awareness and critical AI literacy. Through these kinds of activities AI becomes a catalyst for teaching rhetorical competence, foregrounding the interpretive and interactive dimensions of academic writing rather than reducing it to surface fluency. Clearly, teacher education programmes should include explicit training in understanding the linguistic and rhetorical capacities of LLMs, so that instructors are better prepared to guide students in their use. Such training could include genre awareness, stance construction, prompt engineering, and critical discourse analysis of AI-generated texts.

The framework also has important implications for ***assessment***. Current rubrics that privilege clarity, coherence, and grammatical accuracy risk

inadvertently rewarding AI outputs over student learning. To safeguard construct validity, assessment practices must be recalibrated to emphasise reasoning, rhetorical adaptability, and responsiveness to audience expectations. Embedding process-oriented approaches, including staged drafting, annotated reasoning, and reflective commentary, ensure that assessments capture the intellectual work of students rather than the polished surface fluency of machine-generated text. Transparency in AI use should be encouraged, with students asked to reflect explicitly on how they employed tools such as ChatGPT and what rhetorical limitations they identified. Such practices not only protect academic integrity but also foster critical literacy by making students more aware of their own rhetorical choices and the limitations of algorithmic fluency.

Ethically, the integration of AI into academic writing raises pressing questions of authorship, agency, and accountability. Here institutions must establish clear policies that balance innovation with integrity, guiding students to use AI responsibly and reflectively. Encouraging learners to acknowledge AI's role in their writing processes helps demystify its contribution while reinforcing the value of human rhetorical competence. We need, then, as teachers, to find ways to help students in understanding that while AI can assist in generating ideas and drafting texts, key areas of academic writing such as critical thinking, personal engagement, and creativity remain, for the moment, outside of the abilities of LLMs. In this way, AI can be positioned as a partner in learning rather than a threat to authenticity, supporting the development of reflective writers who understand both the affordances and the constraints of machine assistance.

Finally, the framework points toward several directions for ***future research***. Longitudinal studies are needed to track how evolving AI models develop rhetorical capacities over time and how students' engagement with them changes. Cross-cultural investigations can explore how AI-generated texts align with diverse academic traditions and rhetorical expectations, while classroom-based studies will be particularly valuable in capturing how students interact with AI in practice and how pedagogical interventions shape their development of rhetorical awareness. Lines of inquiry such as these will be essential for refining frameworks of AI integration and ensuring that higher education

remains responsive to technological change. It also contributes to a teaching--learning context that balances technological innovation with an enduring commitment to cultivating academic literacy.

Table 3 summarises these implications, indicating the framework, its supporting arguments and future directions for research into how we might use AI productively and ethically.

Table 3. Reasons for and implications of using ChatGPT as a writing assistant

Strengths	**Limitations**
Fluency and grammatical accuracy	Weak stance and authorial presence
Structural clarity and coherence	Limited engagement with readers
Lexical sophistication	Formulaic cohesion and bundle use
Accessible models of organisation	Lack of rhetorical depth and nuance
Pedagogical Uses	**Ethical & Assessment Implications**
Contrastive analysis tasks	Risk of privileging AI outputs if rubrics
Revising AI drafts to add stance	overemphasise clarity and coherence
Prompt design experimentation	Need to recalibrate descriptors to value
Support rhetorical awareness	reasoning, adaptability, and audience
Developing critical AI literacy	Combine with process-oriented assessment
	Transparency in AI use
Emerging research Directions	
Longitudinal tracking of AI's evolving rhetorical capacities	
Integration of AI literacy into writing pedagogy	
Positioning AI as scaffold, not surrogate author	

Source: Author's own elaboration.

Conclusions

This study has explored the role of generative AI in academic writing, situating ChatGPT within the broader debates about fluency, rhetoric, and integrity in higher education. I have highlighted both the promise and the risks of AI integration and summarised the results of five recent corpus-based studies by Jiang and Hyland which provide empirical evidence of the rhetorical gaps that distinguish AI-generated texts from student essays. Across stance, cohesion, engagement, argumentation, and genre awareness, the findings underscore a consistent pattern: showing how ChatGPT produces structurally polished texts but lacks the rhetorical depth and interpersonal nuance that define effective academic literacy.

Following a line from this comparative corpus evidence to the classroom, I have proposed a conceptual framework that treats AI outputs as objects of analysis, revision, and reflection. Taking this approach seriously means that writing teachers can harness AI to strengthen what they do best: assisting learners to produce fluent, clear and accessible texts. The framework therefore addresses current limitations of ChatGPT in creating rhetorically competent essays while also emphasising the need to rethink assessment practices to more transparently reveal AI use and to cultivate critical AI literacy, prompting students not only to use AI responsibly but also to recognise and compensate for its rhetorical shortcomings.

Beyond pedagogy, institutions must guide students in navigating questions of authorship and accountability, while scholars should continue to investigate how AI evolves and how learners engage with it across contexts. Ultimately, this paper is a small contribution to the growing work seeking to reframe AI not as a threat to academic integrity but as a catalyst for pedagogical innovation. By embedding AI within a critically informed framework, higher education can prepare students to thrive in hybrid writing ecologies where human creativity and algorithmic fluency intersect. In doing so, we safeguard the intellectual and rhetorical dimensions of academic writing while embracing the opportunities of a digital age. The challenge, then, is not whether AI belongs in the classroom, but how we design practices that ensure students remain the authors of their own learning.

References

Adeshola, I., & Adepoju, A. P. (2023). The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments*, *32*, 1–14. DOI: 10.1080/10494820.2023.2253858.

Aull, L. L., & Lancaster, Z. (2014). Linguistic markers of stance in early and advanced academic writing: A corpus-based comparison. *Written Communication*, *31*(2), 151–183.DOI: 10.1177/0741088314527055.

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, *57*, 100745. DOI: 10.1016/j.asw.2023.100745.

Chegg (2025). Global Student Survey. Chegg, Inc. Retrieved from https://www.chegg.org/global-student-survey-2025. Accessed 12 May 2026.

Deep, P. D., & Chen, Y. (2025). The role of AI in academic writing: Impacts on writing skills, critical thinking, and integrity in higher education. *Societies*, 15(9), 247. DOI: 10.3390/soc15090247.

Gabrielatos, C. (2018). Keyness: Appropriate metrics and practical issues. In: C. Taylor & A. Marchi (Eds). *Corpus approaches to discourse: A critical review* (pp. 225–258). Routledge.

Granger, S. (2013). The Louvain Corpus of Native English Essays (LOCNESS). Louvain-la-Neuve: Université Catholique de Louvain.

Gao, R., Yu, D., Gao, B., Hua, H., Hui, Z., Gao, J., & Yin, C. (2025). Legal regulation of AI-assisted academic writing: Challenges, frameworks, and pathways. *Frontiers in Artificial Intelligence*, *8*, 1546064. DOI: 10.3389/frai.2025.1546064.

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

Ho, V., & Li, C. (2018). The use of metadiscourse and persuasion: An analysis of first year university students' timed argumentative essays. *Journal of English for Academic Purposes*, *33*, 53–68. DOI: 10.1016/j.jeap.2018.02.001.

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. London: Continuum.

Hyland, K. (2026). *Teaching and researching writing* (5th edition). Routledge.

Hyland, K., & Jiang, F. (2016). Change of Attitude? A Diachronic Study of Stance. *Written Communication*, *33*(3), 251–274. DOI: 10.1177/0741088316650399.

Hyland, K., & Jiang, F. (2018). Academic lexical bundles: How are they changing? *International Journal of Corpus Linguistics*, *23*(4), 383–407. DOI: 10.1075/ijcl.17080.hyl.

Hyland, K., Jiang, K., Wang, W. (2022). Metadiscourse across languages and genres: An overview. *Lingua*, *265*, 103205. DOI: 10.1016/j.lingua.2021.103205.

Hyland, K., & Jiang, F. (2023). Interaction in written texts: A bibliometric study of published research. *Studies in Second Language Learning and Teaching*, *13*(4), 903–924. DOI: 10.14746/ssllt.40220.

Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, *15*(4), ep464. DOI: 10.30935/cedtech/13605.

Jiang, F. & Hyland, K. (2024). Does ChatGPT argue like students? Bundles in argumentative essays. *Applied Linguistics*, *20*, 1–17.

Jiang, F. & Hyland, K. (2025a). Rhetorical distinctions: Comparing metadiscourse in essays by ChatGPT and students. *English for Specific Purposes*, *79*, 17–29. DOI: 10.1016/j.esp.2025.03.001.

Jiang, F. & Hyland, K. (2025b). Does ChatGPT write like a student? Engagement markers in argumentative essays. *Written Communication*, *42*(3), 463–492. DOI: 10.1177/07410883251328311.

Jiang, F. & Hyland, K. (2026). Taking a stance: AI generated vs student written argumentative essays. *Applied Linguistics Review*. DOI: 10.1515/applirev-2025-0188.

Jiang, F. & Hyland, K. (in press). AI, argumentation and assessment: An analysis of interactive metadiscourse in ChatGPT and student writing. *Journal of Pragmatics*.

Jiang, F., & Ma, X. (2018). 'As we can see': Reader engagement in PhD candidature confirmation reports. *Journal of English for Academic Purposes*, *35*, 1–15. DOI: 10.1016/j.jeap.2018.05.003.

Kelly, R. (2024). *Survey: 86% of Students Already Use AI in Their Studies*. Campus Technology. Retrieved from https://campustechnology.com/articles/2024/08/28/survey-86-of-students-already-use-ai-in-their-studies.aspx. Accessed 12 May 2026.

Kim, J., Lee, S., Detrick, R., Wang, J., & Li, N. (2025). Students–Generative AI interaction patterns and their impact on academic writing. *Journal of Computing in Higher Education*, *38*, 504–525. DOI: 10.1007/s12528-025-09444-6.

Li, H., & Wu, X. (2025). The use of generative AI tools in academic writing: A systematic review of research trends and thematic insights. *AI Ethics*, *5*, 5821–5840. DOI: 10.1007/s43681-025-00827-0.

Markey, B., Brown, D. W., Laudenbach, M., & Kohler, A. (2024). Dense and disconnected: Analyzing the sedimented style of ChatGPT-generated text at scale. *Written Communication*, *41*(4), 571–600. DOI: 10.1177/07410883241263528.

Pack, A., & Maloney, J. (2023). Using Generative Artificial Intelligence for Language Education Research: Insights from Using OpenAI's ChatGPT. *TESOL Quarterly*, *57*(4), 1571–1582. DOI: 10.1002/tesq.3253.

Rayson, P. (2016). Log-likelihood and effect size calculator. Lancaster University. Retrieved from http://ucrel.lancs.ac.uk/llwizard.html. Accessed 12 May 2026.

Siddiqui, M. N., Pea, R., & Subramonyam, H. (2025). AI in the Writing Process: How Purposeful AI Support Fosters Student Writing. *arXiv:2506.20595*. DOI: 10.48550/arXiv.2506.20595.

Speri, A. (2026). *I wish I could push ChatGPT off a cliff': professors scramble to save critical thinking in an age of AI*. The Guardian. Retrieved from https://www.theguardian.com/technology/ng-interactive/2026/mar/10/ai-impact-professors-students-learning. Accessed 12 May 2026.

Vee, A., Laquintano, T., & Schnitzler, C. (Eds.) (2023). *TextGenEd: Teaching with text generation technologies*. The WAC Clearinghouse. DOI: 10.37514/TWR-J.2023.2982.

Wolfram, S. (2023). *What Is ChatGPT Doing … and Why Does It Work?* Writings. Retrieved from https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/. Accessed 22 May 2026.

Yang, S., Chen, S., Zhu, H., Lin, J., & Wang, X. (2024). A comparative study of thematic choices and thematic progression patterns in human-written and AI-generated texts. *System*, *126*(9), 103494. DOI: 10.1016/j.system.2024.103494.

Zou, H. & Hyland, K. (2020). "Think about how fascinating this is": Engagement in academic blogs across disciplines. *Journal of English for Academic Purposes*, *43*, DOI: 10.1016/j.jeap.2019.100809.

DOI: 10.2478/doc-2026-0008

This is an open access article licensed under the Creative Commons AttributionNonCommercial-NoDerivatives 4.0 International (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Rahmad Nuthihar

Department of Indonesian Language Education, Faculty of Language and Literature Education, Universitas Pendidikan Indonesia, Indonesia

r.nutihar@upi.edu

ORCID ID: 0000-0001-8060-2113

Dadang S. Anshori

Department of Indonesian Language Education, Faculty of Language and Literature Education, Universitas Pendidikan Indonesia, Indonesia

dadanganshori@upi.edu

ORCID ID: 0000-0002-0109-1181

Isah Cahyani

Department of Indonesian Language Education, Faculty of Language and Literature Education, Universitas Pendidikan Indonesia, Indonesia

isahcahyani@upi.edu

ORCID ID: 0000-0001-5060-8469

Andoyo Sastromiharjo

Department of Indonesian Language Education, Faculty of Language and Literature Education, Universitas Pendidikan Indonesia, Indonesia

andoyo@upi.edu

ORCID ID: 0000-0002-2711-0328

Making Students "Safe" Writers with the SAFE Model in the Age of AI

Article history:

Received **16 January 2026**

Revised **8 February 2026**

Accepted **8 February 2026**

Available online **23 June 2026**

Abstract: University students can develop opinion-writing skills through strong argumentative reasoning, while technological advancements should support high-quality text production without relying too heavily on AI text-generating tools. This study investigates students' responses regarding the challenges and needs in writing argumentation-based opinion texts. Data were collected through a questionnaire administered to 97 students enrolled in the Indonesian Language Education Study Program in West Java, Indonesia. Data analysis employed percentage calculations based on the 6+1 Traits of Writing framework, encompassing ideas, voice, word choice, sentence fluency, conventions, and presentation. Argumentative reasoning was examined using the framework of claim, evidence, and reasoning. The findings reveal that students experience the greatest difficulty in writing conventions, while weaknesses in argumentative reasoning are primarily related to the presentation of evidence. To address these challenges, this study integrates the 6+1 Traits of Writing with an argumentation--based reasoning framework to develop the SAFE writing instruction model, consisting of Surveying, Arguing, Framing-Writing, and Editing. The SAFE model utilizes artificial intelligence as scaffolding in the writing process and responds to debates on text ownership in Generative Artificial Intelligence by positioning students as the primary agents of text production.

Keywords: argumentation, Generative Artificial Intelligence (GenAI), opinion writing, SAFE model, text ownership

Introduction

The writing of arguments for popular articles or opinion pieces can be facilitated by artificial intelligence (AI) chatbots, including ChatGPT, Gemini, DeepSeek, and other text-generation tools (Aydin et al., 2025; Hysaj et al., 2024; Salvagno et al., 2023). However, the ease of writing afforded by AI does not necessarily yield texts rich in argumentation. The language produced by AI often appears unnatural, indicating that argumentative writing still requires the writer's voice to convey perspectives meaningfully in response to particular events (Hinton & Wagemans, 2022). This is because opinion articles constitute a form of scholarly writing that articulates the author's viewpoint on contemporary issues. AI systems are not yet capable of performing this task effectively, as limitations in accessing or interpreting up-to-date events require writers to elaborate their arguments independently. Nevertheless, AI can still serve as scaffolding, assisting writers in organizing ideas that can subsequently be developed in detail within the text (Li & Wilson, 2025).

For university students in Indonesia, particularly those enrolled in Indonesian Language Education programs, writing popular articles is a compulsory course that requires students to produce opinion texts. In completing such tasks, AI serves as a supportive tool that facilitates students' writing processes (Tseng & Warschauer, 2023). However, the use of AI without a clear structure and a sound framework for opinion writing may become counterproductive for students (Guo et al., 2025). Arguments in AI-assisted texts tend to be weak and show limited authorial voice. As a result, opinion pieces often read like summaries of facts or others' views rather than presenting a clear argumentative stance by the writer. Although AI can produce texts with near-perfect adherence to conventions, its word choice often remains limited. Recurrent patterns can also be identified in the writing, particularly in concluding sentences that signal closure, which tend to follow similar formulaic structures (AlAfnan & MohdZuki,

2023). While AI functions as a tool to assist users, the prompts provided must be grounded in a solid conceptual understanding of opinion writing (Ranade et al., 2025). Effective prompts should integrate an appropriate writing model with a strong argumentative framework to ensure that the resulting texts demonstrate high quality and substantive argumentation (Budiharso, 2017).

Writing models in the twenty-first century have continued to evolve, and their application has increasingly been adapted as structured prompts for use with artificial intelligence (AI) systems (Mezzadri & Paita, 2025). Fundamentally, writing models originate from a process-oriented approach to writing (Patthey--Chavez et al., 2004). Within this framework, writing is commonly classified into several stages, namely prewriting, drafting, revising, editing, and publishing (Williams, 2014, pp. 106–107). While these stages emphasize writing as a process, they do not sufficiently address the contextual and content-related dimensions of a text. To address this limitation, Spandel (1997) developed the 6 Traits of Writing model, which consists of the components of ideas, organization, voice, word choice, sentence fluency, and conventions. This model was later refined by the Northwest Regional Educational Laboratory (NWREL) by adding a sixth trait, presentation, resulting in the 6+1 Traits Writing model. The 6+1 Traits Writing framework became a registered trademark of NWREL—now known as Northwest Education—and has been widely recognized for its effectiveness in supporting students' writing development. Empirical studies have demonstrated that the 6+1 Traits Writing model functions both as an evaluative tool and as an instructional framework that effectively enhances students' writing skills (Culham, 2003).

The 6+1 Traits of Writing have been empirically proven to be effective in improving writing skills. This evidence motivated the present study to integrate the seven traits with the argumentative reasoning framework proposed by Sampson et al. (2011). Accordingly, this study addresses the following Research questions:

1) What needs and challenges do university students face in writing opinion articles based on the 6+1 Traits of Writing combined with an argumentative reasoning framework?

2) What affirmative opinion-writing instructional model can be developed for AI-assisted writing by integrating the 6+1 Traits of Writing with an argumentative reasoning framework?

To address the research problems, this article is supported by a literature review that discusses the 6+1 Traits of Writing, the argumentation framework proposed by Sampson and Clark (2008), and the use of artificial intelligence in opinion writing. Furthermore, a descriptive qualitative methodology is employed to facilitate the synthesis of findings related to students' needs and challenges in writing opinion texts. Based on these findings, the study proposes a writing model that utilizes artificial intelligence as scaffolding to support the development of arguments in opinion writing. These findings provide a foundation for developing an AI-supported opinion-writing model. In principle, artificial intelligence is not merely an assistive tool that diminishes users' argumentative abilities, but rather functions as cognitive scaffolding that stimulates critical reasoning, strengthens argument construction, and supports the writing process without displacing the central role of the human writer.

Literature Review

The 6+1 Traits of Writing

According to Coe et al. (2011), the 6 Traits of Writing were developed in the 1980s as a classroom-based assessment approach intended to provide a clearer structure for both teachers and students in understanding what constitutes effective writing (Culham, 2003). The approach also draws on elements of the writing process, including the recursive use of planning, drafting, evaluating, and revising, as developed by Flower and Hayes (1981). The 6+1 Traits of Writing are integrated with stages of the writing process. The model emphasizes six core characteristics of effective writing: ideas and content, organization, word choice, sentence fluency, conventions, and voice, along with an additional trait, presentation. Each trait aligns with writing process stages, including prewriting,

drafting, revising, editing, and publishing. During prewriting, writers identify purpose, audience, and form, and engage in brainstorming. Drafting focuses on developing central ideas and organizing them coherently. Revising attends to the flow of ideas and sentence variety, while editing emphasizes corrections in spelling, grammar, and mechanics. The publishing stage involves presenting the final text clearly and effectively to the intended audience. The model helps writers understand that writing is a recursive process that requires systematic thinking, structuring, and refining ideas.

The 6+1 Traits of Writing can be implemented across age levels and learning contexts. Culham (2003) developed practical guidelines for applying the model in schools in her book *The 6+1 Traits of Writing: The Complete Guide for Grades 3 and Up.* The model is not only applicable in formal instruction but also recommended by educators for use in writing training programs, as reported in an evaluation study by Werkmeister (2010). Nevertheless, research on this model remains limited, partly because the traits themselves are inherent components of writing, leading educators to favor more familiar instructional models such as mind mapping, brainstorming, or roundtable activities. The 6+1 Traits of Writing model shifts the focus of instruction from traditional approaches to the development of writing skills and the provision of meaningful feedback (Kalsum, 2020). It provides explicit criteria that help students improve writing proficiency and critical thinking by clarifying the rationale behind the scores they receive (Culham, 2003).

We found that previous studies on the 6+1 Traits of Writing model consistently demonstrate its effectiveness in improving students' writing skills. Un-udom (2020) found that the 6+1 Traits of Writing model had a positive impact on the writing performance of Thai EFL students, as evidenced by a significant improvement between pre-test and post-test scores ($p = 0.00 < 0.05$) and students' attitudes categorized at a high level ($M = 4.18$). These results align with the study conducted by Qoura and Zahran (2018), reporting that the 6+1 Traits of Writing model was effective in developing students' critical thinking and writing skills through an emphasis on the writing process, rubric-based assessment, and sustained practice. In addition, Werkmeister (2010) concluded that 71% of teachers perceived 6+1 Traits of Writing training as effective as an initial step, while the remaining 29%

considered it less effective due to the need for more time, deeper engagement, and more focused training formats. Overall, previous research suggests that the 6+1 Traits of Writing model enhances writing skills and is suitable for use as a writing training model. However, the integration of the 6+1 Traits of Writing model with argumentation has not yet been explored. Therefore, this article seeks to develop an adaptive, contemporary writing model that employs artificial intelligence as a supportive scaffold during the writing process.

Sampson and Clark's Argumentation Framework

This research employs the scientific argumentation framework developed by Clark and Sampson (2008) and Sampson and Clark (2008). According to Sampson and Clark (2008), scientific arguments consist of three core components: explanation, evidence, and reasoning. Explanation functions as the answer to the research question and presents a claim that must be evaluated. Evidence consists of observations or measurements that support or refute the claim, while reasoning links the evidence to the explanation by demonstrating why the evidence is valid as a basis for support. The quality of an argument is then assessed through empirical criteria, such as the relevance and sufficiency of the evidence, and theoretical criteria, including completeness, usefulness, and consistency of the explanation with established theory.

The scientific argumentation framework, adapted from Sampson and Clark (2008) in the study *The Impact of Collaboration on the Outcomes of Scientific Argumentation*, explains how scientific arguments are constructed through three components: explanation, evidence, and reasoning. Explanation presents the claim or proposition to be examined and must align with evidence that is logically justified through reasoning. Evidence may include trends over time, differences between groups, or relationships among variables, and is used to support or challenge the explanation. Reasoning justifies, explaining how and why the evidence supports the explanation, and why it can be considered a valid basis for the claim.

Sampson & Clark (2008) further clarify that argument quality is evaluated using two groups of criteria: empirical and theoretical. Empirical criteria

include fit with evidence, sufficiency of evidence, predictive power, and quality of evidence. Additionally, theoretical criteria include the sufficiency, usefulness, and consistency of the explanation with other accepted ideas. These considerations indicate that the quality of a scientific argument is determined not only by the strength of its empirical evidence but also by the precision and coherence of its theoretical explanation. Effective argument evaluation requires balancing valid evidence with reasoning that is logically coherent and theoretically justified.

Sampson and Clark's argumentation framework, also known as Argument-Driven Inquiry (ADI), focuses on examining the strength of scientific arguments in both oral and written forms (Sampson et al., 2011). The ADI model has been widely applied in science disciplines such as biology and chemistry to assess the quality of students' argumentation. The three components of Sampson and Clark's argumentation framework, namely claim, evidence, and reasoning, are not merely features of scientific argumentation but constitute the fundamental foundation of effective argument construction. These components are also more accessible for novice writers than Toulmin's argumentation model, which is more detailed and includes six elements, namely claim, data, warrant, backing, rebuttal, and qualifier (Stofiana et al., 2025). Existing studies predominantly employ argumentation frameworks to analyze texts rather than to design writing instruction models. The development of argumentation-based models as a foundation for writing instruction remains limited. However, integrating an argumentation framework with a writing model has the potential to produce a robust and effective formula for writing instruction. Therefore, this study integrates Sampson and Clark's argumentation framework into a writing model that employs artificial intelligence as scaffolding.

The Use of AI in Opinion Writing

Artificial intelligence (AI) can generate texts, including argumentative texts, in accordance with user instructions (Macagno, 2021; Vassiliades et al., 2021). In higher education writing instruction, AI has become a tool that facilitates

students' idea generation processes (Črček & Patekar, 2023). However, AI exhibits notable limitations, particularly in voice, as generated texts often appear rigid and overly formal (Jiang & Hyland, 2025). This contrasts with student-produced writing, which tends to demonstrate a stronger authorial voice supported by arguments grounded in the most up-to-date factual information. Despite its limitations in conveying authorial voice, AI should not be viewed as detrimental to writing instruction, as instructors play a central role in integrating AI effectively into writing pedagogy. In practice, AI-supported writing often relies on tools such as ChatGPT and Grammarly. However, their application in developing argumentative writing tends to be limited, functioning primarily as tools for linguistic editing (Stofiana et al., 2025). However, through appropriate utilization, AI has the potential to support the production of texts that approximate the quality of professional writing.

Writing opinion articles for the media is a form of public-oriented communication that requires alignment with readers' needs and expectations (Anshori, 2023, p. 203). The development of opinion writing skills, especially for novice writers, can be facilitated through the application of effective and systematic writing steps. Writing may be conceptualized as a staged process that involves selecting a theme or idea, preparing relevant references, identifying the target audience, determining an appropriate development strategy, formulating an engaging title, composing an introduction, developing the main body, articulating a clear stance, and constructing a coherent conclusion (Abed, 2024; Flower & Hayes, 1981). Structurally, an opinion article typically consists of a title, lead, body, and conclusion, with each component adapted to comply with the word limits and editorial standards of the target media outlet.

University students often encounter multiple challenges in academic writing, including fear of starting, uncertainty about when to begin, difficulties with organization, and linguistic limitations (Mammadova, 2023). These challenges align with findings from previous research, which identify difficulties such as generating ideas, organizing thoughts, and dealing with limited information literacy (Hyland, 2019; Tardy, 2016).

Method

Research Design

This study employed a descriptive quantitative design with a survey approach to analyse students' needs regarding an opinion writing model based on the 6+1 Traits of Writing and argumentative reasoning. This approach was selected because it aligns with the Research objective, which aims to systematically map students' perceptions and needs.

Participants

The participants consisted of 97 undergraduate students of Indonesian Language Education from public and private universities in West Java Province. The participants were selected using purposive sampling based on the following criteria: (1) currently active as undergraduate students, (2) have completed a course in opinion writing, and (3) voluntarily agreed to participate. The distribution of participants by gender and semester is presented in Table 1.

Table 1. Distribution of Participants

N		**97**
Gender	Female	84
	Male	13
Semester	3	30
	4	38
	5	29

Source: Authors' own elaboration.

Instruments and Data Collection Procedures

The research instrument consisted of an online questionnaire developed using Google Forms. The questionnaire comprised three main sections. The first section collected demographic information. The second section included twenty closed-ended statements using a five-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) to measure students' needs related to the components of the 6+1 Traits of Writing (fourteen items) and argumentative reasoning (six items). The third section consisted of five open-ended questions designed to explore challenges and learning needs in opinion writing in greater depth. The development of the questionnaire was based on the theoretical framework of the 6+1 Traits of Writing proposed by Culham (2003) and the argumentation framework by Sampson and Clark (2008), with two items designed for each examined aspect.

Data collection was conducted from October to November 2025 using a survey method, with questionnaires distributed through class WhatsApp groups facilitated by lecturers teaching the opinion writing course. Student participation in this study was voluntary, and participants were free to withdraw from the questionnaire at any time without any academic consequences. In the opening section of the questionnaire, participants were required to provide informed consent before proceeding to items exploring challenges and needs related to opinion writing. All data provided by participants were treated confidentially and used solely for research purposes. No personal identifying information is reported in this article, and participants are identified using the code "R" (respondent) followed by a numerical sequence corresponding to the order of questionnaire completion.

Data Analysis

Data analysis was conducted by integrating quantitative and qualitative datasets. Quantitative data from the twenty closed-ended items were analyzed using descriptive statistics, including the calculation of means and standard deviations to describe the overall tendencies of student responses. The percentage of

agreement levels was then computed using an index formula with interpretation intervals of 0 to 19.99 percent (Strongly Disagree), 20 to 39.99 percent (Disagree), 40 to 59.99 percent (Neutral), 60 to 79.99 percent (Agree), and 80 to 100 percent (Strongly Agree).

Meanwhile, qualitative data from the five open-ended questions were analyzed using thematic analysis through a systematic process that included: (1) data transcription and familiarization, (2) generating initial codes based on patterns in student responses, (3) identifying and categorizing themes, (4) reviewing and refining themes for consistency, and (5) reporting the findings supported with representative excerpts from student responses. The integration of quantitative and qualitative findings provided a comprehensive understanding of students' needs in learning to write opinions.

A single researcher conducted data analysis in this study; therefore, inter-rater agreement was not applied. This approach is consistent with the interpretative nature of linguistic analysis (Willis & Harvey, 2025). To ensure trustworthiness, the study employed four key strategies: (1) prolonged engagement through the researcher's sustained involvement in data collection and analysis; (2) triangulation by comparing data obtained from multiple data collection techniques; (3) an audit trail through systematic documentation of all stages of the research process; and (4) peer debriefing through discussions with colleagues to examine the consistency and accuracy of the analytical findings (Nowell et al., 2017).

Findings

The 6+1 Traits of Writing as Student Needs

Based on the 14-item measure of the 6+1 Traits of Writing, 97 students provided 1358 responses. The responses consisted of 23 Strongly Disagree, 72 Disagree, 329 Less Agree, 511 Agree, and 423 Strongly Agree. The total score obtained was 5313. The highest possible score for Strongly Agree was 5 multiplied by

1358, which equals 6790. The lowest possible score for Strongly Disagree was 1 multiplied by 1358, which equals 1358. Using the index formula, the resulting score was 78.25 percent, which falls into the Agree category. These results indicate that the 6+1 Traits of Writing are perceived as highly necessary for students in opinion writing.

A descriptive analysis of the 14 items assessing students' perceived needs for the application of the 6+1 Traits of Writing in opinion writing showed consistently high results. The combined mean score for the seven traits was 3.85 (SD = 0.95), reflecting a strong overall level of agreement. These findings suggest that students view the 6+1 Traits of Writing as highly relevant and essential for developing their competence in opinion writing. Further analysis of each trait revealed variation in perceived need levels, as summarized in Table 2. The two traits with the highest mean scores were Presentation (M = 4.09, SD = 0.98) and Conventions (M = 4.03, SD = 0.98), both of which fall within the very high category. Conversely, Organization received the lowest mean score (M = 3.72, SD = 0.95), although it still falls within the high category. Descriptive statistics of students' needs regarding the 6+1 Traits of Writing can be seen in Table 2.

Table 2. Descriptive Statistics of Student Needs on the 6+1 Traits of Writing

Traits	Mean	SD	Interpretation
Presentation	4.09	0.98	Very High
Conventions	4.03	0.98	Very High
Voice	3.84	0.95	High
Ideas	3.77	0.90	High
Word Choice	3.75	0.92	High
Sentence Fluency	3.75	0.99	High
Organization	3.72	0.95	High
Overall	**3.85**	**0.95**	**High**

Source: Authors' own elaboration.

The high perceived need for Presentation and Conventions indicates strong student awareness of the importance of technical and formal aspects of writing. This may reflect their academic experiences with formal writing standards that emphasize mechanical accuracy, formatting, and layout. The relatively lower score for Organization suggests that students face challenges in constructing coherent, logically structured texts. These challenges include the ability to systematically arrange ideas, establish clear relationships between paragraphs, and build an effective argumentative flow. Although the aspects of Ideas (M = 3.77) and Word Choice (M = 3.75) fall within the high category, these scores indicate the need for deeper development of idea elaboration and lexical precision. The findings for Voice (M = 3.84) indicate that students wish to express their unique identity and perspective in writing, though this expression may not yet be fully developed or consistent.

Overall, these results reveal a profile of students' needs, showing their confidence in and priority for the technical and mechanical components of writing (Presentation and Conventions). In contrast, they place comparatively less emphasis on components that require higher-order thinking and complex reasoning (Organization and Ideas).

Needs for Argumentative Reasoning Skills

Six items were administered to assess students' needs regarding the application of argumentative reasoning in opinion writing. 97 students completed the six items with a total of 580 responses. The responses consisted of 14 Strongly Disagree, 34 Disagree, 134 Less Agree, 251 Agree, and 147 Strongly Agree. The total score obtained from these responses was 2223. The highest possible score for Strongly Agree was 5 multiplied by 580, for a total of 2900. The lowest possible score for Strongly Disagree was 1 multiplied by 580, which equals 580. The index formula yielded a score of 76.66%, placing it in the Agree category. These findings demonstrate that the argumentative reasoning framework by Sampson and Clark is perceived as necessary by students in opinion writing.

Analysis of the six items measuring students' needs in argumentative reasoning revealed high levels of agreement. The combined mean score for the three components of argumentation was 3.78 (SD = 0.96), indicating strong awareness of the importance of argumentative structures in opinion writing. As shown in Table 3, the component with the highest mean score was Evidence (M = 4.00, SD = 0.98), followed by Reasoning (M = 3.68, SD = 0.97) and Claim (M= 3.66, SD = 0.93). The students' needs for argumentative reasoning components are shown in Table 3.

Table 3. Students' Needs for Argumentative Reasoning Components

Components	Mean	SD	Category
Evidence	4.00	0.98	Very High
Reasoning	3.68	0.97	High
Claim	3.66	0.93	High
Overall	**3.78**	**0.96**	High

Source: Authors' own elaboration.

The high Evidence score reflects students' strong understanding of the importance of supporting arguments with data, facts, and credible sources. This aligns with their academic experience, where the use of references is a standard expectation in scholarly writing. However, the lower scores for Claim and Reasoning highlight areas requiring additional support. Difficulties formulating clear, assertive claims suggest that students still struggle to articulate their argumentative positions explicitly. Meanwhile, the Reasoning component, which also falls within the high category, indicates the need for students to improve their ability to build logical relationships between evidence and claims and to evaluate the strength of their own arguments.

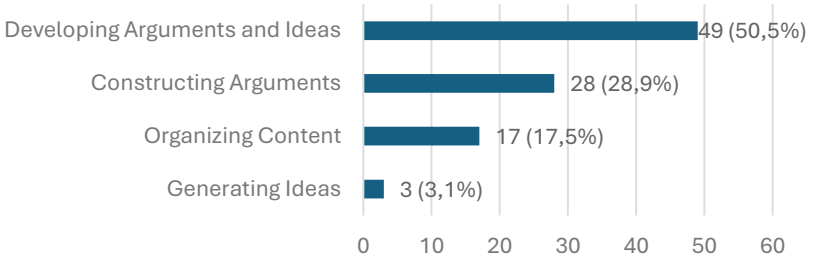
Overall, these findings indicate a gap between students' ability to gather evidence and their ability to formulate clear argumentative statements and construct coherent chains of reasoning. This discrepancy highlights the need

for instructional approaches that not only emphasize evidence gathering but also strengthen students' skills in claim formulation and logical reasoning.

The Most Difficult Aspects of Writing an Opinion

Students' responses regarding the most difficult aspects of writing an opinion were classified into four categories, namely (1) developing arguments and ideas, (2) constructing arguments, (3) organizing content, and (4) generating ideas. In these four categories, each respondent provided only one primary answer, and the total prevalence is 100 percent. This is shown in Figure 1.

Figure 1. The Most Difficult Aspects of Writing an Opinion



Source: Authors' own elaboration.

Figure 1 shows that the most difficult aspect of writing an opinion is developing arguments and ideas, with a prevalence and distribution of 50.5 percent. The second most difficult aspect is constructing arguments, with a prevalence and distribution of 28.9 percent. Meanwhile, the lowest category is generating ideas, with a prevalence and distribution of 3.1 percent. Based on these responses, it can be concluded that most students reported difficulties in developing arguments and ideas.

Writing an opinion requires the writer to develop arguments and ideas so that the resulting text exerts a strong Influence on readers. One response from R (55) indicated that difficulty in developing arguments occurs because it requires logical reasoning and involves critical thinking skills, as shown in excerpt 1.

1) "In my opinion, the most difficult part of writing an opinion is constructing an argument. The reason is that constructing an argument does not only require personal views, but must also be supported by logical reasoning, strong evidence, and concrete examples to convince readers. Sometimes, ideas are already present, and the content is well organized, but it is difficult to formulate reasons that are truly relevant and accountable. In addition, when constructing an argument, we also need to consider different reader perspectives so that the writing does not appear one-sided. Therefore, this stage requires critical thinking skills and the ability to choose appropriate words so that the opinion is conveyed clearly and persuasively". (R-55)

The difficulty experienced by R-55 is similar to other responses. The key terms related to challenges in developing arguments and ideas found in students' answers include "evidence", "logical", and "critical thinking". This can be seen in several responses presented in excerpts 2, 3, and 4.

2) "In my opinion, the most difficult part is constructing an argument. This is because it requires critical thinking skills to connect opinions with relevant evidence, and to write with a logical flow so that readers can easily understand the intention". (R-47)
3) "The most difficult part for me is constructing an argument. This is because I need to ensure every opinion is supported by logical reasoning and evidence, and arranged in a convincing flow so that readers can understand my perspective". (R-12)
4) "In my opinion, the most difficult part of writing an opinion is constructing an argument. This is because constructing an argument requires critical and logical thinking so that the opinion expressed is not only..." (R-18)

Based on the four categories of difficulties students experience in writing opinions, each can be further summarized into five main points. The explanation

of these categories and key points related to difficulties in writing opinions is presented in Table 4.

Table 4. Summary Points of Opinion Writing Difficulties by Category

No.	Category of Difficulties	Main Summary Points
1.	Developing Arguments and Ideas	a) Difficulty in constructing a logical and systematic framework of thought b) Challenges in connecting personal ideas with relevant supporting data or facts c) Concerns about the ability to persuade readers with strong reasoning d) Need for concrete evidence and examples to support the opinion e) Difficulty in developing initial ideas into complete and structured arguments
2.	Constructing Arguments	a) Difficulty in articulating ideas into effective and persuasive sentences b) Challenges in selecting precise words to convey opinions c) Concerns about the alignment of arguments with proper writing conventions d) Need for critical thinking skills when constructing reasons e) Difficulty in balancing personal opinions with objective facts
3.	Organizing Content	a) Difficulty in constructing a coherent and easily understandable writing structure b) Challenges in organizing the flow from introduction to body to conclusion c) Concerns about coherence between paragraphs and sentences d) Need for a clear writing framework without losing the main idea e) Difficulty in maintaining focus to keep writing on the main topic
4.	Generating Ideas	a) Difficulty in identifying topics that are interesting, relevant, and current b) Challenges in selecting an original and meaningful point of view c) Concerns about the feasibility of developing an idea into a full opinion piece d) Need for broad insight and adequate references e) Difficulty in determining an appropriate issue that aligns with personal interests and knowledge

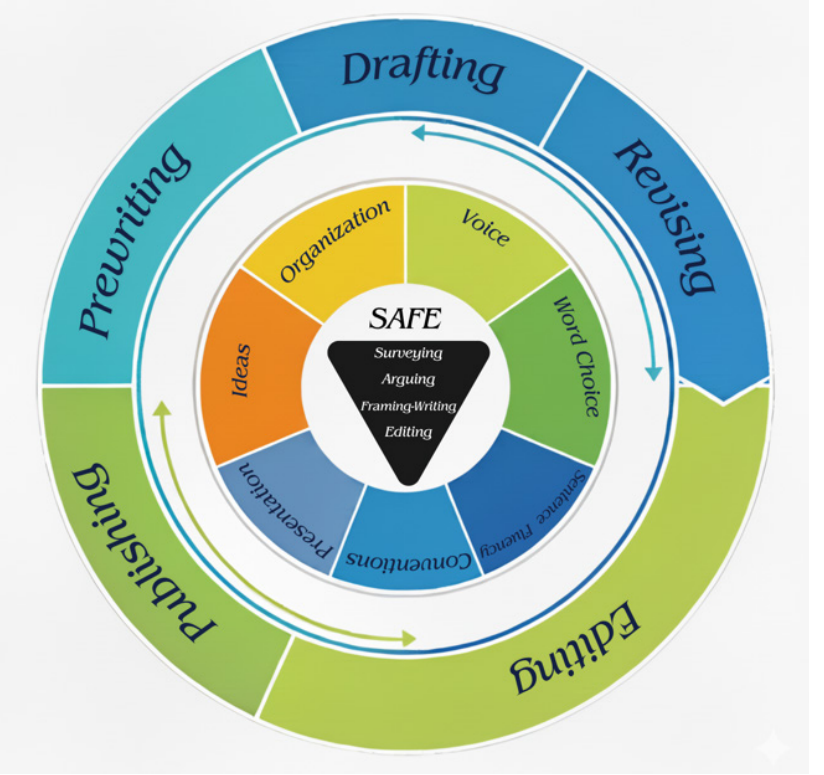
Source: Authors' own elaboration.

Based on Table 4, it can be concluded that, in developing arguments and ideas, the primary obstacle lies in constructing a logical, systematic framework and bridging the gap between subjective ideas and relevant empirical support. In the dimension of argument construction, students face challenges in transforming abstract thoughts into effective sentences that adhere to linguistic conventions while maintaining a balance between personal perspectives and factual objectivity. The aspect of content organization reveals difficulties in building structural coherence, including establishing a systematic flow from introduction to conclusion and sustaining focus without topic deviation. Meanwhile, in the stage of idea selection, the main issues revolve around identifying current topics that are original and sufficiently promising to be developed into substantive opinion discourse, which requires comprehensive insight and adequate references.

Discussion

The findings indicate that the university students in this study have not yet fully developed proficiency in opinion writing. The primary challenge identified lies in writing conventions, while, in terms of argumentative reasoning, students experience significant difficulty presenting evidence. These challenges prompted the proposal of an opinion-writing model that integrates the 6+1 Traits of Writing with an argumentative reasoning framework (Clark & Sampson, 2008, 2007; Sampson et al., 2011). This proposed writing model, termed SAFE—an acronym for Surveying, Arguing, Framing-Writing, and Editing—is also grounded in a process-oriented approach to writing (Coffin et al., 2003, p. 34). Such an approach is essential to ensure that writers clearly understand the stages involved in constructing opinion texts. A schematic representation of the SAFE model is presented in Figure 2.

Figure 2. A Diagram of the SAFE Model for Opinion Writing

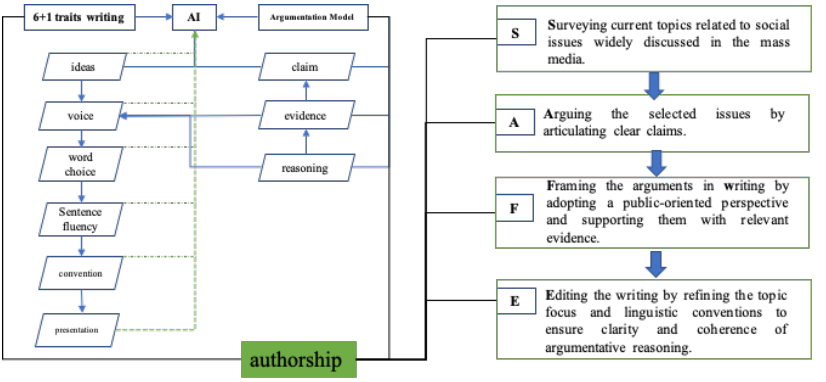


Source: Authors' own elaboration.

The SAFE model is designed to be particularly effective and easily applicable for novice writers through the strategic use of AI. The authorship of opinion texts remains with the writer, as the articles are based on the writer's own ideas. In principle, AI operates solely based on user-generated instructions; therefore, the quality of the resulting opinion text depends on the clarity and substance of the prompts and inputs provided by the writer. The tension between human authorship and AI assistance can be addressed through the SAFE model. SAFE functions as a syntactic framework in which AI is positioned as a mediating tool between the 6+1 Traits of Writing and the argumentative reasoning framework. Argumentative reasoning informs and shapes the application of the 6+1 Traits

of Writing. At the same time, AI is used solely as scaffolding to support writers in exploring ideas, selecting words, and ensuring grammatical accuracy. Control over the writing process remains entirely in human hands. A detailed explanation of the SAFE model's hypothetical design is illustrated in Figure 3.

Figure 3. Inputs and Processes of the SAFE Model for Opinion Writing



Source: Authors' own elaboration.

Based on Figure 3, the SAFE model's inputs are fundamentally divided into two components: the 6+1 Traits of Writing model and the argumentative reasoning model. The SAFE model originates from an academic claim grounded in current and relevant events. This claim is closely connected to the idea's component of the 6+1 Traits of Writing. At this stage, AI can be used to support idea exploration, enabling writers to develop an initial conceptual framework for composing opinion texts. Furthermore, evidence and reasoning are associated with voice, which constitutes a defining characteristic of argumentative writing. Voice cannot be replaced by any level of sophistication in text-generating AI. It remains inherently human, as each writer possesses a distinctive way of articulating arguments, while AI functions solely as scaffolding in this process. The remaining traits, including word choice, sentence fluency, conventions, and presentation, can be refined through the use of AI tools. For instance, applications such as Grammarly can be employed to check grammatical accuracy and suggest

more appropriate lexical choices. Grammarly also offers thesaurus features that enable greater lexical variation. Nevertheless, AI exhibits limitations in selecting vocabulary that is fully relevant to specific disciplinary domains and argumentative contexts. Therefore, human intervention remains essential to ensure that word choice aligns appropriately with the subject matter and the intended argumentative stance.

The SAFE model explicitly positions artificial intelligence as a supportive tool in the process of opinion writing rather than as a substitute for the writer's critical and argumentative thinking. This conceptual stance underscores that the primary strength of opinion writing continues to rest on human capacity to construct claims, select relevant evidence, and formulate reasoning grounded in personal experience, knowledge, and perspective (Akinwande et al., 2024; Bouziane & Bouziane, 2024). Furthermore, voice is an essential dimension of argumentative writing that cannot be replicated by AI, as it is inherently subjective, contextual, and ideological (Lehman, 2026; Lehman & Sułkowski, 2021). This perspective is reinforced by previous research demonstrating that digital technologies do not replace the human role in meaning-making and the formation of argumentative stance, as both processes are closely tied to the writer's critical awareness and ideological positioning (Alt & Kapshuk, 2022; Deep & Chen, 2025).

One of the primary difficulties identified among students in this study concerns the conventions aspect of writing, which can be effectively addressed through the use of AI. The capability of AI to detect and correct spelling errors is well established. However, users must possess sufficient prior schematic knowledge of orthographic rules to make informed decisions when applying AI-generated corrections. In the Indonesian context, orthography is regulated by the Decree of the Head of the Agency for Language Development and Fostering of the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, Number 0424/I/Bs.00.01/2022, concerning the Indonesian Spelling System. This regulation is commonly referred to as *Ejaan yang Disempurnakan (EYD)*. It provides comprehensive guidelines on letter use, word formation,

punctuation, and the spelling of loanwords[1]. The use of artificial intelligence is particularly relevant for addressing the conventions component, as this aspect of writing is mechanical, rule-based, and relatively stable normatively. The existence of standardized spelling and orthographic rules enables AI systems to perform technical corrections consistently, although the interpretation of communicative context and purpose remains beyond the scope of such technology. Therefore, writers' mastery of orthographic conventions remains a fundamental prerequisite for ensuring that AI-generated corrections can be accurately understood and appropriately applied (Yang, 2025). AI thus functions solely as a technical support tool in writing instruction rather than as a substitute for writers' linguistic competence, a position reinforced by digital literacy studies that conceptualize technology as scaffolding in the development of writing skills (Arseven & Bal, 2025; Wang & Wang, 2025).

The SAFE model functions as a syntactic framework for writing. Each stage of this model incorporates the core elements of argumentative reasoning, namely claim, evidence, and reasoning (Sampson et al., 2011). Inputs derived from the 6+1 Traits of Writing and the argumentative reasoning model, which AI scaffolds, are processed through the SAFE model. The surveying stage requires writers to explore current issues that reflect public concerns. Such issues can be identified through mass media sources, particularly online news outlets. The use of AI helps writers map contemporary issues by examining trending or popular topics on mass media platforms. The arguing stage involves selecting a focal issue and formulating a claim in response. This claim serves as a foundation for determining a tentative title, which can later be refined. Writers tend to apply their ideas more effectively when a clear title anchors a text. Title construction based on the formulated claim can also be supported by AI, which can generate alternative title options for further consideration.

---

[1] The EYD is publicly accessible online at https://ejaan.kemendikdasmen.go.id/. However, AI systems such as ChatGPT have not yet been fully updated to reflect the latest version of the Indonesian Spelling System and may still rely on the earlier Ejaan Bahasa Indonesia (EBI) framework, which is no longer in effect.

The SAFE model positions the argumentation process as the core of opinion writing, while artificial intelligence supports the stages of idea exploration and initial formulation. The sophistication and speed of AI in processing large volumes of information enable writers to map issues and generate alternative titles more systematically, without taking over the process of determining argumentative stance (Zhao, 2025). Nevertheless, writers retain a central role in constructing claims and directing the text's overall trajectory as thinking subjects. AI functions solely as scaffolding to enhance the efficiency and organization of the writing process (Cardon et al., 2023). This approach aligns with research on argumentative writing that emphasizes integrating human reasoning and technological support in writing instruction in the digital era (Khampusaen, 2025).

Rejecting the integration of AI in writing instruction means denying the role of technological advancements in education. Such concerns are comparable to earlier resistance to adopting word-processing or reference-management software. In principle, AI is not intended to replace human writers; rather, its presence facilitates the articulation of ideas, particularly in the construction of arguments (Alexander et al., 2023; Sułkowski et al., 2025). In argumentative writing based on the SAFE model, AI is positioned as a pedagogical resource that enhances students' argumentation and writing skills. AI is employed to support the development of argumentative structures grounded in claim, evidence, and reasoning, as well as to provide formative feedback on the 6+1 Traits of Writing. The use of AI is directed as cognitive and linguistic scaffolding rather than as a text-generating agent, thereby ensuring that critical thinking, creativity, and academic responsibility remain fully under students' control.

Conclusion

This study demonstrates that students' difficulties in writing opinion texts are not limited to mechanical aspects of writing or conventions, but also involve weaknesses in argumentative reasoning, particularly in presenting relevant and logical evidence. These findings highlight the importance of helping students

combine basic writing skills with a clear structure for presenting arguments, so they can write opinion texts that are logical and academically sound. Within the context of writing instruction in the digital era, the use of artificial intelligence has the potential to support students; however, its application must be carefully regulated to prevent the displacement of the human writer's role and the emergence of overreliance on technology.

Within process-oriented and argumentation-based writing pedagogy, the SAFE model offers a human–AI hybrid framework that positions AI as cognitive support rather than a text generator. Grounded in the 6+1 Traits of Writing and argumentative reasoning, the model reinforces writers' agency and textual ownership, thereby refining conceptual understandings of authorship in literacy and language education amid the growing use of generative AI. The findings also demonstrate that the SAFE model provides lecturers with an ethical, structured guide for integrating AI into opinion writing instruction, while fostering students' awareness of authorship, argumentative competence, and responsible AI use. In addition, the model is relevant to digital literacy curriculum development and the design of instructional materials that balance technological integration with the development of critical thinking.

Several limitations should be noted, as the research involved respondents from a single study program comprising 97 students. In addition, the study did not conduct an in-depth analysis of the quality of the students' opinion texts. Future research should assess the SAFE model through direct implementation in AI-supported opinion writing and analysis of students' written products across the SAFE stages.

Acknowledgment

The authors thank Lembaga Pengelola Dana Pendidikan (LPDP), The Finance Ministry of Indonesia, for supporting the publication of this article.

References

Abed, T. B. (2024). Implementing the process writing approach to teach paragraph writing at Birzeit University. *Journal of Language Teaching and Research*, *15*(1), 24–34. DOI: 10.17507/jltr.1501.04.

Akinwande, M., Adeliyi, O., & Yussuph, T. (2024). Decoding AI and human authorship: Nuances revealed through NLP and statistical analysis. *International Journal on Cybernetics & Informatics (IJCI)*, *13*(4), 85–103. DOI: 10.5121/ijci.2024.130408.

AlAfnan, M. A., & MohdZuki, S. F. (2023). Do artificial intelligence chatbots have a writing style? An investigation into the stylistic features of ChatGPT-4. *Journal of Artificial Intelligence and Technology*, *3*(3), 85–94. DOI: 10.37965/jait.2023.0267.

Alexander, K., Savvidou, C., & Alexander, C. (2023). Who wrote this essay? Detecting AI-generated writing in second language education in higher education. *Teaching English With Technology*, *23*(2), 25–43. DOI: 10.56297/BUKA4060/XHLD5365.

Alt, D., & Kapshuk, Y. (2022). Argumentation-based learning with digital concept mapping and college students' epistemic beliefs. *Learning Environments Research*, *25*(2), 687–706. DOI: 10.1007/s10984-021-09385-8.

Anshori, D. S. (2023). *Genre bahasa media: Literasi bahasa di panggung media [Media language genres: Language literacy in the media sphere]*. PT Refika Aditama.

Arseven, T., & Bal, M. (2025). Critical literacy in artificial intelligence assisted writing instruction: A systematic review. *Thinking Skills and Creativity*, *57*, 101850. DOI: 10.1016/j.tsc.2025.101850.

Aydin, O., Karaarslan, E., Erenay, F. S., & Bacanin, N. (2025). Generative AI in academic writing: A comparison of DeepSeek, Qwen, ChatGPT, Gemini, Llama, Mistral, and Gemma. *ArXiv Preprint ArXiv:2503.04765*. DOI: 10.48550/arXiv.2503.04765.

Bouziane, K., & Bouziane, A. (2024). AI versus human effectiveness in essay evaluation. *Discover Education*, *3*(1), 201. DOI: 10.1007/s44217-024-00320-6.

Budiharso, T. (2017). Developing prompts of argumentative essay writing for EFL college students. *Indonesian Journal of Language Teaching and Linguistics*, *2*(3), 227–244. DOI: 10.30957/ijoltl.v2i3.531.

Cardon, P., Fleischmann, C., Aritz, J., Logemann, M., & Heidewald, J. (2023). The challenges and opportunities of AI-assisted writing: Developing AI literacy for the AI age. *Business and Professional Communication Quarterly*, *86*(3), 257–295. DOI: 10.1177/23294906231176517.

Clark, D. B., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching*, *45*(3), 293–321. DOI: 10.1002/tea.20216.

Clark, D. B., & Sampson, V. D. (2007). Personally-seeded discussions to scaffold online argumentation. *International Journal of Science Education*, *29*(3), 253–277. DOI: 10.1080/09500690600560944.

Coe, M., Hanita, M., Nishioka, V., & Smiley, R. (2011). *An investigation of the impact of the 6 + 1 trait writing model on grade 5 student writing achievement. Final report.* National Center for Education Evaluation and Regional Assistance.

Coffin, C., Curry, M., Goodman, S., Hewings, A., Lillis, T., & Swann, J. (2003). *Teaching academic writing: A toolkit for higher education*. Routledge.

Črček, N., & Patekar, J. (2023). Writing with AI: University students' use of ChatGPT. *Journal of Language & Education*, *9*(4), 128–138. DOI: 10.17323/jle.2023.17379.

Culham, R. (2003). *6 + 1 traits of writing: The complete guide grades 3 and up*. Northwest Regional Educational Laboratory.

Deep, P. Das, & Chen, Y. (2025). The role of AI in academic writing: Impacts on writing skills, critical thinking, and integrity in higher education. *Societies*, *15*(9), 1–19. DOI: 10.3390/soc15090247.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*(4), 365–387. DOI: 10.2307/356600.

Guo, K., Zhang, E. Di, Li, D., & Yu, S. (2025). Using AI-supported peer review to enhance feedback literacy: An investigation of students' revision of feedback on peers' essays. *British Journal of Educational Technology*, *56*(4), 1612–1639. DOI: 10.1111/bjet.13540.

Hinton, M., & Wagemans, J. H. M. (2022). How persuasive is AI-generated argumentation? An analysis of the quality of an argumentative text produced by the GPT-3 AI text generator. *Argument & Computation*, *14*(1), 59–74. DOI: 10.3233/AAC-210026.

Hyland, K. (2019). *Second language writing*. Cambridge University Press.

Hysaj, A., Freeman, M., & Hamam, D. (2024). Using AI tools to enhance academic writing and maintain academic integrity. In: A. Coman, S. Vasilache (Eds.), *Social Computing and Social Media. HCII 2024. Lecture Notes in Computer Science* (pp. 57–66), vol 14704. Springer, Cham. DOI: 10.1007/978-3-031--61305-0_4.

Jiang, F. (Kevin), & Hyland, K. (2025). Does ChatGPT argue like students? Bundles in argumentative essays. *Applied Linguistics*, *46*(3), 375–391. DOI: 10.1093/applin/amae052.

Kalsum (2020). The 6+1 trait writing model to build up students' writing. *Eduvelop: Journal of English Education and Development*, 4(1), 23–30. DOI: 10.31605/eduvelop.v4i1.799.

Khampusaen, D. (2025). The impact of ChatGPT on academic writing skills and knowledge: An investigation of its use in argumentative essays. *LEARN Journal: Language Education and Acquisition Research Network*, *18*(1), 963–988. DOI: 10.70730/PGCQ9242.

Lehman, I. M. (2026). Voice and identity negotiation in academic publishing: A writing pedagogy for EAL doctoral students. *TESOL Quarterly*. DOI: 10.1002/tesq.70065.

Lehman, I. M., & Sułkowski, Ł. (2021). Representation of voice in English essays of non-native students of business. *Innovations in Education and Teaching International*, *58*(2), 168–181. DOI: 10.1080/14703297.2020.1712221.

Li, M., & Wilson, J. (2025). AI-integrated scaffolding to enhance agency and creativity in K-12 English language learners: A systematic review. *Information*, *16*(7), 519. DOI: 10.3390/info16070519.

Macagno, F. (2021). Argumentation schemes in AI: A literature review. Introduction to the special issue. *Argument & Computation*, *12*(3), 287–302. DOI: 10.3233/AAC-210020.

Mammadova, T. (2023). Academic writing and information literacy instruction in digital environments: A complementary approach. In: *Academic Writing and Information Literacy Instruction in Digital Environments: A Complementary Approach*. Palgrave Macmillan.

Mezzadri, M., & Paita, M. (2025). Artificial intelligence in language education: Exploring prompting strategies to foster argumentative writing skills. *Journal of Studies in Language, Culture and Society (JSLCS)*, *8*(1), 188–205.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, *16*(1), 1–13. DOI: 10.1177/1609406917733847.

Patthey-Chavez, G. G., Matsumura, L. C., & Valdés, R. (2004). Investigating the process approach to writing instruction in urban middle schools. *Journal of Adolescent & Adult Literacy*, *47*(6), 462–476.

Qoura, A. A., & Zahran, F. A. (2018). The effect of the 6+1 trait writing model on ESP university students critical thinking and writing achievement. *English Language Teaching*, *11*(9), 68–79. DOI: 10.5539/elt.v11n9p68.

Ranade, N., Saravia, M., & Johri, A. (2025). Using rhetorical strategies to design prompts: A human-in-the-loop approach to make AI useful. *AI & SOCIETY*, *40*(2), 711–732. DOI: 10.1007/s00146-024-01905-3.

Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Can artificial intelligence help for scientific writing? *Critical Care*, *27*(1), 75. DOI: 10.1186/s13054-023--04380-2.

Sampson, V., & Clark, D. (2008). The impact of collaboration on the outcomes of scientific argumentation. *Science Education*, *93*(3), 448–484. DOI: 10.1002/sce.20306.

Sampson, V., Grooms, J., & Walker, J. P. (2011). Argument-Driven Inquiry as a way to help students learn how to participate in scientific argumentation and craft written arguments: An exploratory study. *Science Education*, *95*(2), 217–257. DOI: 10.1002/sce.20421.

Spandel, V. (1997). *Dear parent: A handbook for parents of 6-trait writing students*. Northwest Regional Educational Lab.

Stofiana, T., Sunendar, D., Mulyati, Y., & Sastromiharjo, A. (2025). Writing with AI, thinking with Toulmin: metacognitive gaps and the rhetorical limits of argumentation. *Ampersand*, *15*, 100242. DOI: 10.1016/j.amper.2025.100242.

Sułkowski, Ł., Chidlow, A. A., & Lehman, I. M. (2025). Language and generative AI: A new paradigm of organizational research. *Discourses on Culture*, *24*(1), 19–43. DOI: 10.2478/doc-2025-0012.

Tardy, C. M. (2016). *Beyond convention: Genre innovation in academic writing*. University of Michigan Press.

Tseng, W., & Warschauer, M. (2023). AI-writing tools in education: If you can't beat them, join them. *Journal of China Computer-Assisted Language Learning*, *3*(2), 258–262. DOI: 10.1515/jccall-2023-0008.

Un-udom, S. (2020). Effects of the 6+1 traits of writing model on Thai EFL students' writing achievement. *Journal of Research and Development Institute, Rajabhat Maha Sarakham University*, *7*(1), 241–254.

Vassiliades, A., Bassiliades, N., & Patkos, T. (2021). Argumentation and explainable artificial intelligence: A survey. *The Knowledge Engineering Review*, *36*(e5), 1–35. DOI: 10.1017/S0269888921000011.

Wang, C., & Wang, Z. (2025). Investigating L2 writers' critical AI literacy in AI-assisted writing: An APSE model. *Journal of Second Language Writing*, *67*, 101187. DOI: 10.1016/j.jslw.2025.101187.

Werkmeister, B. E. (2010). *Evaluation of 6+1 trait writing and the professional development training used to implement the program*. Lindenwood University.

Williams, J. D. (2014). *Preparing to teach writing: Research, theory, and practice* (4th ed.). Routledge.

Willis, R. K., & Harvey, S. (2025). Doing interpretative phenomenological analysis (IPA) in applied linguistics research: A field-specific guide. *Research Methods in Applied Linguistics*, 4(3), 100279. DOI: 10.1016/j.rmal.2025.100279.

Yang, A. (2025). Generative AI chatbots: The future of grammar and spelling correction in English learning. *International Journal of Information and Communication Technology*, *26*(22), 72–87. DOI: 10.1504/IJICT.2025.146909.

Zhao, D. (2025). The impact of AI-enhanced natural language processing tools on writing proficiency: an analysis of language precision, content summarization, and creative writing facilitation. *Education and Information Technologies*, *30*(6), 8055–8086. DOI: 10.1007/s10639-024-13145-5.

DOI: 10.2478/doc-2026-0009

This is an open access article licensed under the Creative Commons AttributionNonCommercial-NoDerivatives 4.0 International (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Djalel Baghzou

University of Salah Boubnider Constantine, Algeria

djalal.baghzou@univ-constantine3.dz

ORCID ID: 0009-0008-0633-215X

Assala Belsem Bouameur

University of Salah Boubnider Constantine, Algeria

assalabelsem.bouameurunivconstantine3.dz

ORCID ID: 0009-0007-2881-550X

Michał Szostak

Institute for Management Research, Civitas University, Warsaw, Poland

michal.szostak@civitas.edu.pl

ORCID ID: 0000-0002-7774-2964

Mourad Mahmoudi

Emir Abdelkader University of Islamic Sciences, Algeria

m.mahmoudi@univ-emir.dz

ORCID ID: 0009-0003-8351-0400

From Mental Concept to Algorithmic Exemplification: Rethinking the Aesthetic Situation in the Age of Generative AI

Article history:

Received	**7 April 2026**
Revised	**15 April 2026**
Accepted	**22 April 2026**
Available online	**23 June 2026**

Abstract: The emergence of generative artificial intelligence (AI) necessitates a re-evaluation of the foundational structures of artistic creation and reception. This article utilises the "aesthetic situation" theory, originally formulated by Maria Gołaszewska (1984) and recently expanded through a managerial lens by Michał Szostak (2023b), to analyse the transformative impact of computational agents on the creative process. Building upon the distinction between "supporting" and "substituting" AI roles (Szostak & Modliński, 2025), this paper introduces a conceptual model of "algorithmic exemplification". It posits that generative AI functions as a technical mediator, translating an artist's internal "mental concept" into a publicly accessible artefact through iterative prompting and latent-space navigation. By systematically analysing the five components of the aesthetic situation—the artist, the artwork, the recipient, the world of values, and the real world—this article argues that generative AI does not merely automate production but reconfigures the ontic status of the artwork and the agency of the creator. The discussion highlights

the shift from direct craftsmanship to "aesthetic situation management", in which the artist's role evolves into that of a curator and prompt engineer, potentially substituting for traditional manual skills while supporting higher--order conceptual synthesis.

Key words: aesthetic situation, generative artificial intelligence, Maria Gołaszewska, algorithmic exemplification, mental concept, computational creativity

Introduction

The integration of generative artificial intelligence into the arts has moved beyond a technical novelty to a fundamental disruption of aesthetic theory. Traditional frameworks of aesthetics often assume a human-centric creative act, in which the bridge between an internal vision and a physical object is crossed through human labour and technical skill. However, the capacity of Large Language Models (LLMs) and diffusion models to generate high-fidelity artefacts challenges the traditional definitions of authorship and the nature of the artwork itself.

The aesthetic situation, as a structural totality, has historically presupposed a direct, causal link between the artist's cognitive states and the material properties of the artwork. This link was forged through the artist's physical interaction with a medium—whether paint, clay, or musical notation—where the "mental concept" was externalised through a series of motor and technical choices. In the era of generative AI, this causality is disrupted. The artist no longer engages in the manual manipulation of matter but instead interacts with a high-dimensional latent space. This shift introduces a layer of algorithmic mediation that complicates the ontic relationship between the creator's intent and the final artefact, necessitating a radical rethink of how an aesthetic situation is managed and sustained.

Furthermore, the emergence of generative systems forces a reconsideration of the "artistic act" itself. If the creative process is viewed as a series of managerial decisions, as suggested by Szostak (2023b), then the introduction

of AI offers new solutions that either augment human capability or replace it entirely. This "supporting versus substituting" dichotomy (Szostak & Modliński, 2025) is not merely a technical distinction but a philosophical one. It questions whether the "artist" node in Gołaszewska's (1984) model requires a human consciousness to initiate the situation, or whether a human acting as a "prompt manager" is sufficient to maintain the integrity of the aesthetic encounter.

The disruption of the causal chain between the mental concept and its physical manifestation implies a shift in the location of "creativity" within the aesthetic situation. In classical interpretations, creativity was often viewed as a form of poiesis—a bringing-forth that required the artist's direct, hands--on struggle against the material's resistance. However, when an algorithmic interface replaces the medium, the "resistance" shifts from the physical to the linguistic and the statistical. The artist must navigate the probabilities of a pre-trained model, making the creative act an exercise in "reflexivity" and selection (Szostak, 2023a). This suggests that the essence of the artistic act in the age of AI is no longer found in the execution, but in the precision of the initial mental concept and the critical evaluation of the algorithmic output.

Moreover, this reconfiguration suggests a new ontology for the artwork, moving away from the "intentional object" described by Ingarden (1981) toward what might be termed a "computational by-product" that requires human re--contextualisation to achieve aesthetic status. If the generative model lacks intentionality, the responsibility for the "world of values" (Gołaszewska, 1986) rests entirely on the human agent's ability to manage the situation. The implications are profound: if the AI substitutes the creator's technical role, the human artist must double down on their role as a moral and aesthetic gatekeeper to prevent the "devaluation" of the art market (Fortuna & Modliński, 2021). The "real world" context thus demands greater transparency into the level of algorithmic mediation to preserve the social trust inherent in the artist--recipient relationship.

The transition from a purely human-driven process to one mediated by algorithms also affects the "world of values" surrounding the artwork. Values such as originality, authenticity, and genius have long been anchored in the scarcity of human talent and the uniqueness of human effort. When

generative AI can produce infinite variations of an aesthetic style, the value of the resulting "algorithmic exemplification" must be re-evaluated. This does not necessarily lead to a devaluation of art, but rather a shift in value toward the conceptual and reflexive stages of the creative process—the initial mental concept and the final curation.

Moreover, the recipient's role is fundamentally altered when the artwork's provenance is non-human. The aesthetic situation is a communicative event; when the "sender" of the aesthetic message is a machine, the recipient's capacity for empathy and interpretation is challenged. As Fortuna and Modliński (2021) note, the perception of an AI as a "counterfeiter" rather than a creator can significantly alter the market and cultural reception of the work. This suggests that the "real world" component of the aesthetic situation is increasingly sensitive to the transparency of the algorithmic mediation used during the creative process.

Ultimately, this article addresses the problem of how generative AI alters the structural dynamics of the "aesthetic situation". The central inquiry focuses on the transition from the "mental concept"—the artist's internal, subjective intention—to the "algorithmic exemplification", defined here as the computationally generated manifestation of that intent. By utilising the framework established by Gołaszewska (1984) and refined by Szostak and Modliński (2025), this study explores whether AI serves as a support mechanism for human creativity or a substitute for the human creator within the aesthetic ecosystem.

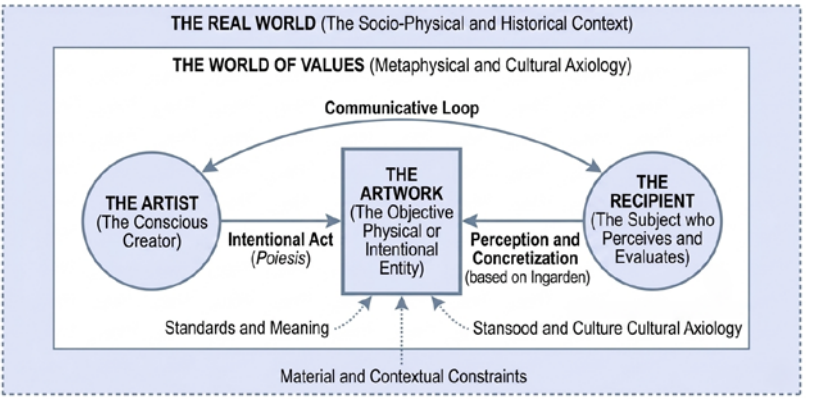
Theoretical Framework

The Concept of Aesthetic Situation

The foundation of this analysis is the "aesthetic situation", a concept developed by Maria Gołaszewska (1984, 1986), who posited that aesthetics should not focus solely on the artwork or the observer in isolation but on a complex, dynamic system comprising five essential elements: (1) The Artist: The conscious creator who initiates the process; (2) The Artwork: The objective physical or intentional entity;

(3) The Recipient: The subject who perceives and evaluates the work; (4) The World of Values: The metaphysical or cultural context providing meaning; and (5) The Real World: The socio-physical environment in which the situation occurs.

Figure 1. The structure of the aesthetic situation according to Gołaszewska (1984)



Source: Own elaboration based on Gołaszewska (1984).

Gołaszewska's approach was influenced by Roman Ingarden's (1970, 1981) phenomenology, emphasising that the artwork is an "intentional object" that requires the recipient's participation to realise its aesthetic qualities fully. Within this framework, the aesthetic situation is not a static arrangement but a processual event. The "Artist" does not merely produce an object; they encode a specific set of values derived from the "World of Values" into a material substrate. The "Recipient" then decodes these values, but this decoding is always situated within the "Real World"—the physical, social, and historical conditions that dictate how an object is perceived. The logical coherence of this system depends on the "Artist" maintaining a primary, intentional role as the source of the aesthetic impulse.

Building upon this, the relationship between the "Artist" and the "Artwork" is traditionally one of direct ontic dependence. The artwork exists because the artist willed it into being through a specific technical act. In Gołaszewska's

(1984) view, this act is saturated with the artist's personality and their unique grasp of the "World of Values". Consequently, the "Artwork" serves as a bridge, a material crystallisation of a spiritual or conceptual intent. If the intervention of a non-conscious agent fractures this bridge, the entire structural integrity of the aesthetic situation is called into question, as the "Recipient" may no longer be engaging with a human consciousness, but with a statistical facsimile of one.

The Aesthetic Situation Management

Szostak (2023b) extended this theory by introducing the concept of "aesthetic situation management". In this view, the creative process is not only an inspired act but a managed sequence of decisions. Szostak and Modliński (2025) further refined this by categorising AI's role into two distinct modes: supporting and substituting. In the supporting role, AI acts as a tool (e.g., software for refining textures), leaving the primary agency to the human. In the substituting role, AI takes over core creative decisions (e.g., generating a complete composition from a brief prompt), potentially displacing the human from the traditional "artist" node in the aesthetic situation.

The shift toward "management" implies that the artist's virtuosity is no longer confined to manual dexterity or technical execution. Instead, the artist acts as a meta-creator, overseeing the interactions among the five elements of the situation (Szostak, 2023b). When AI enters this system in a supporting capacity, it enhances the artist's ability to manifest their "mental concept" by reducing technical friction. In this modality, the intentional link between the artist and the artwork remains robust; the AI is simply an extension of the artist's "real world" tools. The logical structure of Gołaszewska's (1984) model remains intact, with the human creator serving as the definitive "conscious initiator" of the situation.

However, the substituting role of AI introduces a structural tension within the aesthetic situation. If an algorithmic system generates the formal qualities of the "Artwork" based on probabilistic patterns rather than a subjective "mental concept", the causal path from "Artist" to "Artwork" becomes fragmented. In this scenario, the AI performs functions traditionally reserved for the human

creator—such as selecting colour palettes, harmonic progressions, or linguistic structures—thereby assuming a quasi-authorial position. This substitution challenges the "World of Values" (Gołaszewska, 1986), as the recipient may no longer be interacting with a human's intentional expression, but with a "counterfeit" of creativity (Fortuna & Modliński, 2021). The management of the situation thus shifts from content creation to the validation of algorithmic outputs.

This substitutive capacity forces a logical re-evaluation of the "Artist" node. If the AI generates the work, the human agent's role moves from poiesis (making) to proairesis (choosing). The artist becomes a manager of possibilities within a latent space, where the "mental concept" serves as a filtering mechanism for algorithmic outputs. This suggests that the "Artist" in the age of generative AI is defined less by their ability to generate form and more by their "reflexivity"—the capacity to critically evaluate whether an algorithmic exemplification aligns with a specific "World of Values" (Szostak, 2023a). Without this reflexive oversight, the aesthetic situation loses its teleological direction, becoming merely a by-product of data processing.

Furthermore, the "Real World" implications of this substitution affect the "Recipient's" trust. As noted by Fortuna, Modliński, and McNeill (2022), the perception of the creator matters significantly in how art is priced and valued. If the "Recipient" perceives the "Artist" as having been substituted by a machine, the aesthetic situation may undergo a "devaluation", as the work is seen as lacking the "cyborg" or human labour that typically justifies its status as high art (Fortuna & Modliński, 2021). This necessitates a more transparent form of "aesthetic situation management", where the artist clearly defines the boundaries of AI intervention to maintain the situation's authenticity.

The "World of Values" also undergoes a transformation when AI serves as a substitute agent. Generative models operate by identifying and replicating existing patterns in cultural data. Consequently, they are inherently conservative, drawing from a "World of Values" that is already documented and digitised. This raises the question of whether a substituting AI can ever truly introduce new value into the aesthetic situation, or merely reshuffle existing values (Akerkar, 2019). The human artist's role, therefore, becomes even more critical in injecting

genuine novelty and ethical grounding into a process that might otherwise result in "stylistic stagnation".

Finally, the logical extension of Szostak and Modliński's (2025) distinction suggests that the "Artwork" itself becomes a dynamic variable. In a supporting role, the artwork is a finished, static object. In a substituting role, the artwork is often part of an iterative loop—an "algorithmic exemplification" that is constantly refined by the artist-manager. This fluidity requires the "Recipient" to engage with the work not as a finality, but as a specific instance of a broader computational potential. The aesthetic situation, therefore, shifts from the contemplation of a fixed object to the appreciation of a managed process, where the "Real World" constraints of time and computing power become part of the aesthetic experience itself.

Ultimately, the integration of AI into Gołaszewska's (1984) model does not delete any of the five components but redefines their internal hierarchy. The "Artist" must now manage the "Real World" technicalities of AI while simultaneously anchoring the "Artwork" in a human "World of Values". This dual responsibility highlights that while AI can substitute for the labour of art, it cannot substitute for the responsibility of art. The aesthetic situation in the age of generative AI is thus characterised by a heightened need for "reflexive management" (Szostak, 2023a) to ensure that the algorithmic exemplification remains a true reflection of a human mental concept.

From Mental Concept to Algorithmic Exemplification: A Conceptual Model

This article proposes a new conceptual model to describe the creative act in the age of generative AI. This model centres on two poles: the mental concept and the algorithmic exemplification.

Mental concept is defined as the pre-reflective, subjective aesthetic intention held by the human agent. It includes the emotional state, the desired stylistic parameters, and the work's teleological goal. In traditional art, the mental concept is transformed into an artwork through the artist's physical mastery of a medium.

Algorithmic exemplification refers to the artefact produced by an AI system in response to human input. It is "algorithmic" because its formal properties are determined by probabilistic patterns learned from training data (Cetinic & She, 2022). It is an "exemplification" because it provides a concrete, perceptible instance of the abstract parameters defined by the user.

The transition from the mental concept to the algorithmic exemplification represents a fundamental shift in the "technicity" of the aesthetic situation. In conventional artistic production, the "Artist" (Gołaszewska, 1984) serves as the primary engine of transduction, converting an internal, non-material vision into a material reality through the exertion of physical skill. In this new model, however, the artist's "mental concept" does not interface directly with the medium. Instead, it must be translated into a symbolic linguistic or visual code (the prompt), which then triggers the generative model's navigation of its latent space. This introduces a "semantic gap" in which the artist's original intent may be altered, expanded, or constrained by the algorithm's internal logic.

Logical reasoning suggests that this mediation redefines the "Artwork" from a finished intentional object to a specific selection from a vast field of possibilities. If the mental concept acts as a teleological guide, the algorithmic exemplification serves as a feedback mechanism. The artist provides a prompt, observes the algorithm's output, and iteratively refines the mental concept in response. This "dialogic" relationship implies that the mental concept is no longer a static blueprint but a dynamic state that evolves through its interaction with the AI (Szostak, 2023a). The "Artistic Act" thus becomes an iterative management of this loop, in which the artist seeks to minimise the discrepancy between their internal vision and the algorithmic result.

Furthermore, the algorithmic exemplification possesses an ontic status that is distinct from traditional artistic "mimesis". While a human painter might imitate reality through observation, a generative AI imitates the statistical distribution of existing art (Cetinic & She, 2022). This means that the algorithmic exemplification is essentially a "style-translation" of the mental concept. The artist is not merely creating an object; they are navigating a pre-existing "World of Values" (Gołaszewska, 1986) encoded in the machine's parameters. Consequently, the algorithmic exemplification is always a collaborative output

between the individual human intent and the collective cultural data upon which the AI was trained.

The implications for "Aesthetic Situation Management" (Szostak, 2023b) are significant. The artist must develop a new form of "digital virtuosity" that is not based on motor skills but on the ability to manipulate the latent space effectively. This requires a deep understanding of how specific inputs (prompts, seeds, hyperparameters) correlate with specific aesthetic outputs. If the artist lacks this "algorithmic literacy", the transition from mental concept to exemplification becomes haphazard, and the AI effectively shifts from a "supporting" tool to a "substituting" agent that dictates the final form of the work (Szostak & Modliński, 2025).

Moreover, the algorithmic exemplification introduces the element of "randomness" or "stochasticity" into the aesthetic situation. Unlike the precision of a controlled brushstroke, generative models often produce unexpected artefacts. These "emergent properties" can either enhance or detract from the original mental concept. The artist's role as a manager is to decide whether these unintended algorithmic variations possess aesthetic value. This reinforces Szostak's (2023a) emphasis on "reflexivity"—the artist must be a critical judge of the machine's "creativity", ensuring that the final "Artwork" still resonates with human values and intentions.

This model also highlights a shift in the "Real World" manifestation of art (Gołaszewska, 1984). Because the transition to an algorithmic exemplification is near-instantaneous compared to traditional methods, the "scarcity" of the artistic act is diminished. This speed enables rapid exploration of many mental concepts, potentially leading to a more "concept-heavy" form of art. However, it also poses a risk of "aesthetic saturation", where the market is flooded with high--fidelity exemplifications that lack a profound mental grounding. The stability of the aesthetic situation thus depends on the recipient's ability to recognise the human "mental concept" behind the "algorithmic facade".

Finally, the relationship between the mental concept and the algorithmic exemplification suggests a new form of "cyborgian" creativity (Harbisson, 2019). When the machine seamlessly translates the human intent, the boundary between the "Artist" and the "Tool" blurs. The algorithmic exemplification

is not just a representation of the artist's thought; it is an expansion of it. In this sense, generative AI provides "new solutions supporting the creator" (Szostak & Modliński, 2025) by externalising complex visions that would be technically impossible for a human to execute manually. The aesthetic situation is thus transformed into a collaborative space where human imagination and algorithmic power co-create the final intentional object.

Ultimately, the conceptual model of "From Mental Concept to Algorithmic Exemplification" posits that the value of AI-assisted art lies in the intentionality of the transition. If the process is managed with reflexivity, the resulting artwork remains an authentic component of an aesthetic situation. If the transition is left entirely to the algorithm, the work risks becoming a mere "counterfeit" (Fortuna & Modliński, 2021), a formal shell without the core of human value that Gołaszewska (1986) identified as the heart of aesthetics.

Rethinking the Aesthetic Situation

The introduction of generative AI necessitates a systematic re-analysis of Gołaszewska's (1984) five components. This section deconstructs how algorithmic mediation reconfigures the ontic and functional roles of each element within the aesthetic situation.

The Artist

The artist's role shifts from a "maker" to a "director" or "curator". When AI serves a substituting role (Szostak & Modliński, 2025), the manual technical skill traditionally associated with the artist is diminished. However, the "reflexivity" of the artist (Szostak, 2023a) remains crucial in selecting which AI-generated exemplification possesses aesthetic value. Logically, the artist remains the "conscious initiator" of the situation, but their agency is now expressed through the management of the "semantic gap" between intent and output. The artist must ensure that the "algorithmic exemplification" does not become an autonomous entity that obscures the original "mental concept".

The Artwork

The artwork loses its status as a unique "autographic" trace of human labour. It becomes a manifestation of a "latent space"—a mathematical representation of stylistic and conceptual possibilities (Cetinic & She, 2022). The ontic status of the AI-generated artwork is precarious; it is often perceived as a "counterfeit" of human creativity (Fortuna & Modliński, 2021). Within the framework of "aesthetic situation management", the artwork is no longer a static end-product but a specific, managed instance of a computational process. This shift implies that the "Artwork" node in Gołaszewska's (1984) model must now account for the "virtual" nature of its creation, where the intentional object exists as a co--production between human selection and algorithmic probability.

The Recipient

The recipient now faces the challenge of "anthropomorphising" the machine or questioning the authenticity of their emotional response (Chen & Ibrahim, 2023). Research indicates that recipients often struggle to differentiate between human and AI-generated works, yet they tend to devalue art once its AI origins are revealed (Fortuna et al., 2022; Köbis & Mossink, 2021). The recipient's role in "completing" the intentional object (Ingarden, 1981) becomes more complex, as they must decide whether to attribute meaning to the AI's "choices" or only to the human artist's prompts. This suggests that the recipient's evaluative capacity is increasingly directed toward the process of management rather than the perfection of the form.

The World of Values

The "World of Values" is complicated by algorithmic bias (Roselli et al., 2019). Because AI models are trained on existing human datasets, they may reinforce historical biases rather than generate truly "new" aesthetic values. The value of "originality" is contested when the work is a probabilistic recombination of existing styles. In a substituting AI context, the "World of Values" risks becoming

a closed loop of historical data (Akerkar, 2019). Therefore, the artist's "reflexivity" (Szostak, 2023a) is the only mechanism through which the aesthetic situation can transcend mere repetition and engage with contemporary moral or metaphysical values. The implication is that "aesthetic value" in the age of AI is increasingly tied to the ethical and conceptual depth of the "mental concept" rather than the technical execution of the "algorithmic exemplification".

The Real World

In the "Real World", the aesthetic situation is governed by the art market and institutional frameworks. The substitution of human labour with AI has significant economic implications. Furthermore, the technology enables "situated art education" and new immersive experiences through VR and AR (Kwon & Morrill, 2022). Beyond the immediate concerns of market devaluation, the "semantic gap" (defined here as the loss of direct physical transduction between the artist's mind and the material medium) fundamentally reconfigures the "Real World" component of the aesthetic situation. In traditional craftsmanship, the "Real World" validates the artwork through the scarcity of manual execution and the physical evidence of labour. However, because the transition from a mental concept to an algorithmic exemplification is near--instantaneous and computationally mediated, the "Real World" must pivot toward valuing the transparency of the management process.

Application of the Concept

This shift necessitates that institutional frameworks, such as galleries, auction houses, and legal bodies, develop new protocols for "provenance". These protocols must account for the semantic mapping of prompts and the iterative history of the "reflexive loop", as the artwork's value becomes inextricably linked to how the human intent bridged the algorithmic gap. Consequently, this transition marks a "technological trans-role conflict", in which the social contract of art no longer relies on the "hand" of the creator but on the ethical disclosure of their role as a meta-manager. The "Real World" thus evolves from

a marketplace of physical artefacts into a marketplace of verified intentionality, where the "authenticity" of the work is substantiated by the documented rigour of the artist's algorithmic navigation.

The socio-physical environment of the aesthetic situation is now permeated by the "black box" of the algorithm, creating a new tension between the accessibility of art production and the transparency required for market authenticity.

The logical integration of these shifts suggests that the "Real World" is currently undergoing a "technological trans-role conflict". As AI systems begin to replicate the functions of the human "Artist", the social and professional boundaries of the creative class are blurred. If the "Real World" institutions (museums, galleries, legal frameworks) fail to distinguish between "supporting" and "substituting" AI applications, the traditional hierarchy of the aesthetic situation may collapse. The management of the situation must therefore extend into the public sphere, where the artist justifies the use of algorithmic mediation as a tool for "persuasive creativity" rather than a mere efficiency measure (Szostak, 2022).

Furthermore, the "World of Values" is fundamentally challenged by the "transparency" of the creative process. In the age of AI, the value of an artwork is inextricably linked to the recipient's knowledge of its provenance. If the "algorithmic exemplification" is presented as purely human-made, the aesthetic situation is built on a "counterfeit" foundation (Fortuna & Modliński, 2021), leading to an inevitable rupture in the recipient's trust once the truth is revealed. This implies that "aesthetic situation management" (Szostak, 2023b) now includes a mandatory ethical component: the disclosure of the AI's role to preserve the "intentional object's" integrity within the "World of Values".

Finally, the implications for the "Recipient" involve a shift from passive observation to "active verification". Because the "Real World" is increasingly saturated with high-fidelity, low-intent artefacts, the recipient must exercise a higher degree of critical reflexivity to distinguish between a "mental concept" externalised through AI and a purely algorithmic output lacking human teleology. The aesthetic situation thus becomes a more demanding cognitive event for all participants. The "Artist" must manage the technology with precision, while the "Recipient" must navigate the "World of Values" with increased scepticism, ensuring that the "Artwork" remains a meaningful bridge between two human subjectivities.

Discussion

The conceptual model of transition from the mental concept to algorithmic exemplification provides a necessary framework for deconstructing the contemporary aesthetic situation. However, a deep analytic interrogation of this model reveals latent contradictions and shifting ontological boundaries that challenge traditional axiology. By examining the three analytic strata—ontological, epistemic, and evaluative—this discussion elucidates the broader implications of generative AI as a mediator in art management.

Revisiting the Supporting-Substituting Dichotomy: The Co-creative Entanglement

The distinction between AI's supporting and substituting roles, as proposed by Szostak and Modliński (2025), serves as a vital taxonomic starting point. Yet, the theoretical analysis indicates that generative AI may be facilitating a third, hybrid mode: "co-creative entanglement". In this state, the boundary between the artist's mental concept and the algorithm's latent space becomes porous. Unlike traditional tools that merely support execution, generative models often "hallucinate" or provide divergent outputs that retroactively alter the artist's original intention. This suggests that the mental concept is not a static precursor to the work but a fluid state that emerges through a feedback loop with the machine. On this view, the AI is neither a mere assistant nor a total replacement, but a non-conscious collaborator that reshapes the "Artist" node into a manager of stochastic possibilities.

The Ontological Status of Algorithmic Exemplification versus Ingarden's Stratification

A comparison of algorithmic exemplification with Roman Ingarden's (1970) theory of the stratification of the artwork reveals a fundamental shift in the "intentional object". Ingarden posited that an artwork consists of several layers—material, phonetic, schematic, and intentional—which require

the recipient's "concretisation" to achieve aesthetic existence. Algorithmic exemplification, however, complicates this hierarchy. Because the "material" layer is often a digital transient, and the "schematic" layer is derived from a statistical latent space (Cetinic & She, 2022), the work's intentionality becomes diffused. The theoretical analysis indicates that while a traditional artwork is an intentional object willed by a human, an algorithmic exemplification is a "probabilistic object". This creates a tension where the recipient may attempt to concretise an intention that the algorithm, by its very nature, does not possess.

Destabilisation of the Originator: The Distributed Mental Concept

Generative AI destabilises the artist's role as the sole originator of the mental concept by introducing "distributed creativity". In Gołaszewska's (1984) framework, the artist is the conscious initiator who draws from a personal world of values. In the age of generative AI, the "mental concept" is often prestructured by the model's capabilities and biases. The artist's imagination is arguably "channelled" through the software's linguistic and visual parameters. A corollary of this argument is that the "World of Values" (Gołaszewska, 1986) is no longer solely external to the machine; it is partially embedded within the training data. Consequently, the creative impulse is divided between the human's subjective desire and the machine's statistical probability, leading to a situation in which the "origin" of the aesthetic idea becomes increasingly difficult to locate.

The Simulation versus Embedding of Values

A critical evaluative question arises: can algorithms embed aesthetic values, or do they merely simulate them? Based on the review by Cetinic and She (2022), AI models do not "understand" beauty or meaning but replicate the formal markers associated with those values in human datasets. Therefore, the "World of Values" in an AI-mediated aesthetic situation is a simulated one. The analysis reveals a latent contradiction: if a recipient experiences genuine aesthetic emotion

(Chen & Ibrahim, 2023) from a simulated value, does the "authenticity" of the source matter? Fortuna and Modliński (2021) suggest that the perception of AI as a "counterfeiter" devalues the situation. This indicates that while the algorithm can exemplify the form of a value, it cannot exemplify its moral or historical weight, which remains the sole province of the human manager.

Epistemological Limits and Embodied Intuition

There are significant epistemological limits to the current model, particularly concerning the role of embodied intuition and the unconscious. Gołaszewska (1984) and Ingarden (1981) emphasise the importance of the lived experience in the aesthetic act. Algorithmic exemplification is a purely cognitive and mathematical process that lacks the "physical resistance" of a medium—the tactile feedback of clay or the resistance of a bow on a string. The theoretical analysis indicates that the current model cannot fully account for the loss of "haptic knowledge" during the transition from a mental concept to an output. Future theoretical work should examine whether the "reflexive management" described by Szostak (2023a) can compensate for the absence of the artist's physical presence and the spontaneous, unconscious "accidents" of manual craft.

Implications for the "Real World" and Art Market Integrity

The transition to algorithmic exemplification has profound implications for the "Real World" component of the aesthetic situation. As high-fidelity artefacts become infinitely reproducible and low-cost, the traditional scarcity that underpins the art market is threatened. If the "Artist" node is perceived as having been substituted by an algorithm, the social contract between the creator and the recipient may fracture. The evidence from Fortuna, Modliński, and McNeill (2022) suggests that the "human element" is the primary driver of value. Therefore, aesthetic situation management must evolve into a practice of "transparency management", in which the artist's primary labour is not the creation of the image but the substantiation of the human intent that justifies the image's existence in a world saturated by synthetic media.

Limitations and Future Research

This conceptual analysis is limited by its primary focus on the structural components of the aesthetic situation and may not fully capture the nuance of specific artistic genres, such as performance art or music, where the role of "supporting" AI might differ significantly from the visual arts. Furthermore, the model assumes a certain level of "algorithmic literacy" among artists, which may not be universally present. Future empirical research could examine whether recipients' "concretisation" of AI art differs when they are informed about the specific prompting process versus when they are told the work is purely human-made. Additionally, researchers should investigate the "long-term stylistic drift" that may occur in the "World of Values" if generative models begin to be trained on their own outputs, potentially leading to a recursive narrowing of aesthetic diversity.

Conclusion

Summary of Principal Findings. This article has demonstrated that the integration of generative artificial intelligence into the creative process effects a fundamental transition from manual craftsmanship to a model of "algorithmic exemplification". The primary theoretical yield of this analysis is the identification of a triadic mediation—comprising the artist's mental concept, the generative interface, and the resulting artefact—which reconfigures the "Artist" node in Gołaszewska's (1984) aesthetic situation into a role of reflexive management. The foregoing reconstruction suggests that while AI may substitute for the technical execution of an artwork, it remains a supporting component of human intentionality only when the artist maintains critical oversight of the transition between internal concept and externalised output.

Theoretical Implications. The conceptualisation of the artist as a manager of the aesthetic situation (Szostak, 2023b) has significant implications for the fields of art management, digital aesthetics, and AI ethics. Within art management,

the shift toward algorithmic exemplification necessitates new frameworks for authenticating "human-centric" value in a market increasingly populated by synthetic media. A corollary of this position is that the "World of Values" (Gołaszewska, 1986) must now account for the "semantic gap" introduced by machine learning models, where the preservation of authenticity depends on the transparency of the artist's intervention. In the realm of AI ethics, this analysis reveals a necessary distinction between "simulated creativity" and "intentional expression"; if the AI substitutes the human without a corresponding mental concept, the resulting situation risks being perceived as a "counterfeit" (Fortuna & Modliński, 2021), thereby eroding the communicative trust between the creator and the recipient. Furthermore, the model suggests that digital aesthetics must move beyond the study of the finished artefact to analyse the iterative "prompt-and-select" loop as a distinct site of artistic virtuosity.

Future Research Directions. While the proposed model provides a robust framework for discrete digital artefacts, its generalizability is limited for performative, ephemeral, or embodied AI installations, which may require further adaptation to account for real-time physical interaction or site-specific variables. To build upon this conceptual foundation, future research could address the following inquiries:

1) How does the recipient's awareness of "algorithmic exemplification" specifically alter their "concretisation" of the intentional object (Ingarden, 1970) compared to traditionally produced works?
2) To what extent does the "black-box" nature of generative models constrain or expand the artist's "reflexive management" (Szostak, 2023a) during the evolution of the mental concept?
3) Empirical studies could investigate whether the perceived "aesthetic value" of a work is more strongly correlated with the complexity of the artist's linguistic prompts or the formal perfection of the algorithmic output.

The transformation of the aesthetic situation by generative AI does not signify the obsolescence of the human creator but rather the elevation of

the "mental concept" as the ultimate arbiter of artistic truth. As the technical barriers to exemplification diminish, the artist's capacity to anchor an algorithmic artefact within a coherent world of human values becomes the defining challenge of the contemporary era. This analysis indicates that the future of aesthetics lies not in the competition between human and machine, but in the rigorous management of the movement from mental concept to algorithmic exemplification.

References

Akerkar, R. (2019). *Artificial intelligence for business*. Springer. DOI: 10.1007/978-3-319-97436-1.

Cetinic, E., & She, J. (2022). Understanding and creating art with AI: Review and outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *18*(2), 1–22. DOI: 10.1145/3475799.

Chen, X., & Ibrahim, Z. (2023). A comprehensive study of emotional responses in AI-enhanced interactive installation art. *Sustainability*, *15*(22). DOI: 10.3390/su152215830.

Fortuna, P., & Modliński, A. (2021). A(I)rtist or counterfeiter? Artificial intelligence as (D)evaluating factor on the art market. *The Journal of Arts Management, Law, and Society*, *51*(3), 188–201. DOI: 10.1080/10632921.2021.1887032.

Fortuna, P., Modliński, A., & McNeill, M. (2022). Creators matter: Perception and pricing of art made by human, cyborgs and humanoid robots. *Empirical Studies of the Arts*, *41*(2), 331–351. DOI: 10.1177/02762374221143717.

Gołaszewska, M. (1984). *Zarys estetyki [Outline of aesthetics]*. Państwowe Wydawnictwo Naukowe.

Gołaszewska, M. (1986). *O naturze wartości estetycznych [On the nature of aesthetic values]*. Wydawnictwo Uniwersytetu Jagiellońskiego.

Harbisson, N. (2019). Hearing colors: My life experience as a cyborg. In: X. Pavie (Ed.), *Creativity, imagination and innovation* (pp. 117–126). World Scientific. DOI: 10.1142/9789813273009_0015.

Ingarden, R. (1970). *Studia z estetyki (T. 3) [Studies in aesthetics (Vol. 3)]*. Państwowe Wydawnictwo Naukowe.

Ingarden, R. (1981). *Wykłady i dyskusje z estetyki [Lectures and discussions on aesthetics]*. Państwowe Wydawnictwo Naukowe.

Köbis, N., & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, *114*. DOI: 10.1016/j.chb.2020.106553.

Kwon, H., & Morrill, K. (2022). Virtual reality: Immersive and situated art education with 360-degree cameras, and augmented and virtual reality technology. *Art Education*, *75*(4), 27–32. DOI: 10.1080/00043125.2022.2053458.

Roselli, D., Matthews, J., & Talagala, N. (2019). Managing bias in AI. In: L. Liu & R. White (Eds.), *WWW '19: Companion Proceedings of the 2019 World Wide Web Conference* (pp. 539–544). ACM. DOI: 10.1145/3308560.3317590.

Szostak, M. (2022). Contextual inspiration and motive in persuasive creativity: Lessons from artistic improvisation. *Discourses on Culture*, *17*(1), 101–131. DOI: 10.36145/DoC2022.05.

Szostak, M. (2023a). Reflexivity in aesthetic situation management. *Discourses on Culture*, *20*(1), 171–207. DOI: 10.2478/doc-2023-0015.

Szostak, M. (2023b). *Sztuka zarządzania – zarządzanie sztuką [The art of management – art management]*. Społeczna Akademia Nauk.

Szostak, M., & Modliński, A. (2025). Artificial intelligence in aesthetic situation management: New solutions supporting or substituting an art creator. *Creativity Studies*, *18*(2), 558–569. DOI: 10.3846/cs.2025.20714.

DOI: 10.2478/doc-2026-0010

This is an open access article licensed under the Creative Commons AttributionNonCommercial-NoDerivatives 4.0 International (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Muhammad Erwan Saing

Faculty of Language and Art, Universitas Negeri Surabaya, Surabaya, Indonesia

muhammadsaing@unesa.ac.id

ORCID ID: 0009-0009-4034-4477

Algorithmic Storytelling and Vernacular Folklore: Rewriting the Nenek Pakande Monster in the Age of Generative AI

Article history:

Received	**26 March 2026**
Revised	**28 April 2026**
Accepted	**24 May 2026**
Available online	**23 June 2026**

Abstract: The emergence of generative artificial intelligence has unsettled long-standing assumptions about authorship, originality, and authenticity in literary and cultural production. While debates often focus on academic writing or creative industries in the Global North, less attention has been given to how AI-mediated rewriting affects vernacular narrative traditions. This article examines the Indonesian folklore figure Nenek Pakande through the lens of ecological dramaturgy to explore how algorithmic text generation reshapes narrative voice and cultural memory. Drawing on conceptual analysis and comparative rewriting experiments, the study argues that AI does not merely redistribute authorship but transforms its ontological conditions by introducing non-experiential textual agency. In doing so, it reconfigures originality from inventive subjectivity to probabilistic recombination and reframes authenticity as a question of relational accountability rather than textual purity. By situating folkloric rewriting within the dynamics of algorithmic culture, this article proposes a model of hybrid ecological authorship that acknowledges technological mediation while maintaining ethical responsibility toward cultural and ecological specificity. Rather than positioning AI as either threat or tool, the paper advances a relational framework for understanding writing in the age of distributed cognition.

Keywords: generative AI, algorithmic writing, folklore adaptation, narrative voice, hybrid authorship

Introduction

The rapid emergence of generative artificial intelligence has unsettled foundational assumptions about writing as a human practice. These issues align with earlier theoretical frameworks that challenged the authority of the writer. Roland Barthes (1977) concept of the "death of the author" transformed our understanding of writing, implying that it represents a realm of diverse interpretations rather than being solely defined by a single individual's intent. Subsequently, Michel Foucault (1984) characterized

the "author-function" as a collection of guidelines rather than as a tangible person. More recently, posthumanist perspectives have focused on collective cognition and the relationship between humans and technology (Hayles, 2012; Braidotti, 2013). Nonetheless, despite these theoretical frameworks foreshadowing the weakening of authorial authority, generative AI presents a fundamentally different scenario: the creation of text happens through probabilistic calculations over extensive datasets, generating language devoid of memory, responsibility, or lived experience.

Discussions of AI and writing largely focus on academic integrity, creative industries, and professional communication, overlooking its impact on local oral traditions rooted in memory and environment. Folklore, as a collective and evolving form tied to community and shared memory, offers a critical lens on this shift. This article examines the Indonesian figure Nenek Pakande—often portrayed as a monstrous embodiment of desire, nature, and transgression—to rethink writing in the AI era, highlighting how her fluid, orally transmitted narratives encode communal anxieties about scarcity, excess, and ecological instability.

The growth of generative AI has made this connection more complex. When asked to "rewrite" or "retell" Nenek Pakande, a large language model utilizes worldwide collected story patterns—scary themes, typical story structures, and writing styles learned from its training. The final text might seem logical and culturally relevant, but it often subtly shifts the story towards common dramatic patterns. Local uncertainties become clearer; environmental details are simplified; moral complexities might be reduced to well-known opposites. This scenario illustrates what Ted Striphas (2015) describes as algorithmic culture, where computer-based processes create and standardize types of cultural expression that spread online.

The transformation is not merely stylistic. It touches on the epistemological foundations of narrative transmission. If folklore has historically been preserved by embodied memory and social performance, AI-mediated rewriting adds a form of textual agency that is not tied to a specific location and has intergenerational continuity. According to Kate Crawford (2021), AI systems are integrated in mostly invisible data extraction and power infrastructures. Therefore, changing

text through algorithms is never neutral; it reflects the patterns found in the most common texts and the power structures of language.

Rather than positioning AI as either a threat or a creative force, this article adopts an ecological dramaturgical lens that views writing as the ongoing reconfiguration of relationships among humans, environments, and symbols. It shifts the focus from whether AI can reproduce folklore to how narrative relations are transformed when voice emerges from probabilistic systems rather than collective memory. Drawing on conceptual analysis and rewritten examples, the study advances three key arguments: generative AI redefines authorship from a shared social role into a distributed computational one; it reframes originality as the recombination of existing materials rather than individual invention; and it destabilizes authenticity by obscuring the relational processes that produce narrative voice. In response, the article proposes a hybrid model of authorship in which AI is integrated into the creative ecology without being granted moral autonomy.

By situating Nenek Pakande within contemporary debates on algorithmic writing, the article contributes a vernacular and ecologically grounded perspective, positioning folklore as a diagnostic lens for understanding how AI reshapes cultural memory. It underscores the need to rethink shared responsibility in digitally mediated storytelling and maps an inquiry that moves from theoretical grounding to narrative ecology and the implications of algorithmic rewriting, ultimately reframing tensions between human and algorithm as matters of relational accountability in an evolving cultural landscape. In this article, the term "vernacular" refers to locally embedded storytelling traditions rooted in oral transmission, communal memory, and ecological specificity rather than merely regional or linguistic distinction.

Despite the rapidly expanding scholarship on generative artificial intelligence and writing, most existing studies focus on technical evaluation, educational applications, or ethical risks associated with large language models (Floridi & Chiriatti, 2020; Dwivedi et al., 2023; Kasneci et al., 2023). While these research shed light on the potential and limitations of AI systems, there has been little consideration paid to how algorithmic text generation interacts with vernacular narrative traditions and culturally entrenched storytelling practices.

At the same time, research in folklore studies and cultural memory has long emphasized the collective and adaptive nature of narrative transmission (Bauman, 1984; Finnegan, 1970; Assmann, 2011). These traditions demonstrate that storytelling historically evolved through distributed authorship and iterative transformation. Yet the implications of this insight for contemporary AI-mediated writing remain underexplored.

This article addresses three central questions: (1) how does generative AI reshape relational structures within vernacular storytelling traditions? (2) how does algorithmic rewriting affect ecological dramaturgy and narrative complexity in the case of Nenek Pakande? and (3) how should authorship, originality, and authenticity be reconsidered in the context of hybrid human–algorithmic writing systems?

Theoretical Framework

Large language models (LLMs) generate text through probabilistic prediction, selecting statistically likely word sequences based on patterns learned from large-scale training corpora. Rather than understanding meaning through embodied cognition, these systems operate through pattern recognition and token prediction. Consequently, AI-generated outputs are shaped by the linguistic distributions, narrative conventions, and cultural asymmetries embedded within their training data.

Authorship Beyond the Sovereign Subject

The concept of authorship started to change well before the advent of artificial intelligence. In the twentieth century, literary theorists began to challenge the idea that the author was the only one responsible for creating meaning. Roland Barthes' major intervention reframed writing as a realm of intertextual plurality, claiming that the author is the result of textual circulation rather than the source (Barthes, 1977). From this perspective, understanding arises from the act of reading instead of the writer's intention. Michel Foucault advanced

this argument by describing the "author-function" as a system of discourse that regulates categories, power, and limits on interpretation. Therefore, being an author is not a fixed reality but a concept that has changed over time based on history.

These conceptual advancements hinted at the author's shift away from a central focus, yet they did not eliminate the idea of humans being responsible. Even in theories that challenge structure, the text still stayed connected to human cultural frameworks. The arrival of generative AI introduces a completely new scenario. Large language models (LLMs) generate written content from extensive data collections through chance-based methods, lacking any purpose, physical form, or responsibility. N. Katherine Hayles (2012) argues that thinking in digital spaces is progressively shared between humans and machines. In this situation, writing transforms into a joint venture between living beings and technological systems.

Posthumanism makes things more complicated. Rosi Braidotti states that the posthuman era requires us to rethink what it means to have a self, focusing on connections and physical realities instead of being independent and self-made. From this angle, artificial intelligence doesn't just take over human creativity; it shows the ways in which creativity has always involved relationships with various tools, systems, and structures. However, dispersed agency does not necessarily imply spread accountability. The ethical dilemma is to determine how responsibility works when textual output arises from mechanisms that operate beyond direct human knowledge.

Authorship in folklore is inherently collaborative and iterative, shaped by communal memory and lived experience rather than fixed ownership. While this decentering of the individual parallels algorithmic generation, the two differ fundamentally: folklore emerges from embodied, environmental contexts, whereas generative AI operates through statistical abstraction detached from lived reality. Thus, AI reconfigures authorship from a socially grounded practice into a computationally mediated one, introducing a subtle yet significant shift toward non-experiential textual production.

Algorithmic Culture and Standardization of Expression

To comprehend how artificial intelligence (AI) reshapes narrative production, Ted Striphas (2015) defines algorithmic culture. In this method, algorithms do more than simply manage how culture is experienced; they play a key role in creating cultural styles by affecting what is seen, chosen, and repeated. Systems that suggest content, analyze data for predictions, and models based on machine learning create the setting in which texts are shared and become accepted.

Generative AI takes this logic from distribution to production. Instead of merely picking content, it generates sentences by relying on the patterns found in the data it learned from. According to Kate Crawford (2021), AI systems are deeply connected to systems of data gathering, work, and global power imbalances. The collections of texts used to educate language models do not serve as neutral sources; instead, they showcase dominant language habits and cultural practices. As a result, AI-generated writings may perpetuate hegemonic narrative frameworks while remaining stylistically diverse.

This tendency can lead to narrative flattening. Detailed cultural themes are turned into easily recognizable symbols around the world; confusion is removed to favor clarity; and local traits are reshaped using global storytelling styles. This process is seldom obvious. Rather, it appears subtly through changes in tone, clear morals, and a flow that aligns with widely recognized storytelling rules.

This phenomena does not mean that AI systems are intentionally biased. Rather, it reflects the probabilistic logic through which they operate. Language models prioritize statistically likely sequences, thereby reinforcing patterns that dominate the corpus. In folkloric rewriting, this may produce texts that are grammatically refined yet relationally thinned—maintaining plot while attenuating ecological specificity.

Changing, Reworking, and the Idea of Being Original

The act of changing folk stories has existed for a long time. The theory of adaptation has acknowledged that retelling stories is a valid form of creativity, instead of just copying others. Linda Hutcheon (2006) argues that adaptation

includes both reusing and creating, giving new meaning by shifting contexts. From this perspective, being original comes from rethinking ideas rather than being completely new.

Nevertheless, generative AI makes the idea of adaptation more complex. When a language model rewrites a narrative, it does this without deep understanding. What it produces shows a blend of existing ideas rather than a thoughtful change. This brings up a big question: can random mixing of ideas be seen as real creative adaptation?

Current ideas about originality often refer to the Romantic view of personal expression. However, folklore challenges this idea since its power comes from group tradition, not individual uniqueness. In this way, folklore's way of retelling stories already questions ownership by a single author. The conflict caused by AI arises from the loss of meaningful relationships, not from a lack of creativity. People adapt by considering the situation, the audience, and moral factors, while automated rewriting focuses on what is statistically probable.

Therefore, what we think of as originality during the time of AI might have to be changed to focus on the role of curators. The role of humans changes from being a lone creator to a connector—selecting prompts, assessing results, and deciding which versions keep cultural and ecological significance. This doesn't lessen creativity; instead, it places it within the processes of defining, understanding, and responsibility.

Understanding Ecological Dramaturgy

In this study, ecological dramaturgy is operationalized through three analytical dimensions: spatial configuration, distribution of agency, and ecological symbolism. These dimensions are used to examine how narrative relations shift under algorithmic rewriting.

This article introduces ecological dramaturgy as an interpretive paradigm in order to explore the consequences of AI-mediated rewriting. Traditionally, dramaturgy has concentrated on the way that relationships are organized in a performance, such as how the characters, location, and time interact to generate meaning. Emphasizing geographical, environmental, and material

connections, an ecological expansion highlights the interdependence of human and non-human actors.

In folklore tales, monsters often act as environmental mediators. They stand for worries about imbalances like famine, overconsumption, deforestation, and moral excess. Their presence signals relational disturbance rather than purely moral deviance. Reading Nenek Pakande through ecological dramaturgy therefore shifts attention from plot mechanics to relational configurations: how does the narrative structure human interaction with landscape, scarcity, and communal boundaries?

When AI rewrites such a narrative, it may preserve surface events while subtly reorganizing relational emphasis. Ecological dramaturgy allows us to observe these shifts by posing the question:

- Which connections are emphasized or downplayed?
- How is the space configured?
- Which forms of agency are emphasized or minimized?
- Does the monster remain an ecological mediator, or become a generic villain?

This framework does not reject AI as inherently corrosive. Rather, it offers analytical resources to understand how algorithmic methods change the connections between elements in storytelling environments. It also grounds the proposal of hybrid authorship developed later in the article. If writing is understood as relational configuration rather than isolated production, then ethical responsibility lies in sustaining relational complexity—even when technological systems participate in textual generation.

Synthesis: From Decentered Author to Relational Accountability

Together, these ideas reveal the effects of rewriting done through AI. The poststructuralist theory broke down the belief in the all-powerful author. Posthumanism demonstrated dispersed cognition. Algorithmic culture showcased how computational systems organize expressiveness. Adaptation theory

legitimized rewriting as a creative endeavor. Ecological dramaturgy includes these ideas by placing writing within connections that extend beyond the single person.

Generative AI accelerates the reconfiguration of authorship, introducing new uncertainties in textual production and shifting the central question from whether authorship persists to how responsibility is redefined within hybrid storytelling systems. Applying this framework to Nenek Pakande, the study examines how ecological relations in traditional variants are transformed through algorithmic rewriting, ultimately proposing a relational model of authorship capable of addressing the epistemic and ethical shifts brought by AI.

Methodology

Conceptual Grounding

At the conceptual level, the study situates itself within interdisciplinary debates on authorship, posthuman subjectivity, adaptation theory, and algorithmic culture (Barthes, 1977; Foucault, 1984; Hayles, 2012; Hutcheon, 2006; Striphas, 2015). Instead of thinking of artificial intelligence as just a technological breakthrough, this article regards it as a cultural situation that changes the ways we know things and perceive beauty. Writing is seen not just as the words we produce, but also as a connection within social and technical systems. This viewpoint aligns with important research in AI, which highlights that creating systems are deeply rooted in the structures of data collection, language power dynamics, and unequal relationships (Crawford, 2021). Therefore, any examination of folklore created by AI must take into account the larger cultural settings where these models are developed and used.

Corpus and Narrative Selection

The folkloric material examined in this article centers on Nenek Pakande, a figure appearing in South Sulawesi narrative traditions in Indonesia. The analysis draws upon documented variants circulated in print compilations and oral retellings, focusing on recurring motifs: insatiable hunger, liminal dwelling

spaces (forest, threshold, edge of village), deceptive hospitality, and moral--ecological imbalance.

Rather than emphasizing a single authoritative account, the study examines Nenek Pakande as a story cluster distinguished by thematic repetition and adaptive diversity. This approach reflects the nature of folklore as collective and iterative rather than fixed. The goal is not philological reconstruction but relational mapping—identifying how ecological and social dynamics are structured across variants.

Algorithmic Rewriting Procedure

To examine how generative AI mediates folkloric rewriting, the study conducted controlled prompting experiments using a large language model. Multiple prompts were tested iteratively to observe recurring narrative tendencies rather than isolated outputs. Prompts were formulated to simulate common rewriting scenarios, including:

"...Retell the story of Nenek Pakande as a short horror narrative".
"...Rewrite Nenek Pakande for a global young adult audience".
"...Summarize Nenek Pakande as a moral fable".

Outputs were analyzed qualitatively rather than quantitatively. The emphasis was on identifying shifts in:

- narrative structure,
- moral framing,
- ecological specificity,
- character complexity,
- spatial configuration,
- tonal register.

The analysis focused on relational patterns rather than isolated lexical choices. For example, did the forest remain a morally ambiguous ecological

space, or was it reframed as a generic horror backdrop? Did hunger function as symbolic imbalance, or merely as villainous trait? Was communal memory preserved, or replaced with individualized moral closure?

Significantly, the aim of these tests was not to find out if AI represented traditional stories "correctly". Instead, they act as heuristic devices, illuminating how algorithmic systems rebuild story relationships under various prompting contexts.

Analytical Strategy: Relational Comparison

The study employs close reading as its primary analytical tool. Careful reading, usually associated with examining literature, is still very good for noticing small changes in attitude, stress, and how a story is told. In this context, it functions as a comparative instrument:

- Variant A: Folkloric narrative structure,
- Variant B: AI-generated rewriting,
- Comparative axis: relational density and ecological configuration.

Rather than assessing fidelity, the comparison assesses relationship transformation. This technique reflects the main idea of the article, which suggests that AI questions the concepts of authorship, originality, and authenticity by changing stories instead of directly lying.

This difference is seen through the perspective of ecological dramaturgy. It looks into how control shifts within the story, how the setting carries moral messages, and how different voices are shared. The research applies this approach to find patterns of simplifying narratives, standardizing morals, and matching genres, which are commonly linked to changes made by algorithms.

Constraints and Area of Study

This research does not assert that it covers all AI systems or folk traditions in full. The examples of rewrites are meant to show concepts, but they do not

represent statistical facts. Additionally, the study concentrates on outputs in the English language, which introduces another factor to consider when dealing with materials from Indonesia. These limits are recognized as part of the epistemic conditions under which AI operates: translation, abstraction, and corpus bias are all essential to its operation.

Nonetheless, the methodological combination of conceptual framing and comparative textual analysis allows for a more focused investigation of how generative AI contributes in story transformation. Focusing on changes in relationships rather than just technical aspects, this research brings a culturally relevant viewpoint to larger conversations about writing during the time of algorithm-based creation.

Interpretive Approach

This article uses a conceptual and critical thinking method instead of following a practical experimental design. The research centers on comparing traditional versions of the Nenek Pakande stories with modern AI-assisted rewrites. Instead of evaluating how well the algorithms work in technical ways, the study looks at how generative systems change storytelling styles, relationships, and cultural significance.

The method draws from various areas such as literary theory, folklore studies, and critical studies of AI. These viewpoints help the researcher to understand AI-created stories within wider social and cultural contexts, rather than viewing them as separate technological items.

The study follows three analytical steps. First, it investigates the storytelling environment of the Nenek Pakande folklore tradition, referring to research in oral literature and cultural memory. Next, it looks at how language models create different narrative versions by mixing and matching text patterns. Finally, it considers how these changes impact the ideas of authorship and the truthfulness of the narrative.

The article's interpretive approach aims to highlight larger discussions about authorship and cultural storytelling in spaces influenced by algorithms.

Nenek Pakande as Ecological Narrative

The Monster as a Mediator

Across documented variants, Nenek Pakande is characterized by insatiable appetite. Hunger is not merely physiological but excessive—an appetite that disrupts social equilibrium. She charms, tricks, uses up, or claims the lives of anyone she meets. At first glance, this might seem like a typical moral story: greed is punished, purity is at risk, and peril is hidden in the woods. Yet, this understanding simplifies the story to basic ethics.

Within an ecological dramaturgical frame, hunger signals imbalance rather than simple vice. The figure of Nenek Pakande often emerges at the boundary between village and forest, domestic and wild space. This liminality is significant. The forest, as stated by various Indonesian stories passed down through generations, are not seen as a dangerous or holy area; instead, they are recognized as a site for gathering resources, a location for spiritual experiences, and a space that could be risky. Nenek Pakande occupies this frontier, embodying concerns about exceeding natural boundaries.

Her hunger can thus be interpreted as symbolic excess—overconsumption, scarcity, or social breakdown. She shows up as both a hunter and a sign of trouble. The weakness of the community highlights bigger issues in how people connect with nature. Rather from acting purely as an enemy, she emphasizes relationship imbalance.

Ecological dramaturgy consider: what relationships does the narrative depict? The village depends on forest resources; children move between safety and risk; hospitality can become deception; appetite can threaten communal survival. These tensions are not resolved through abstract moralization alone but through narrative negotiation of space and conduct.

Spatial Configuration and Moral Ambiguity

Unlike standardized fairy tales with fixed narrative arcs, variants of Nenek Pakande often preserve ambiguity. The monster may be cunning but also pitiable;

frightening yet strangely familiar. In other retellings, she appears dressed as an elderly woman, implying fragility and threat. The ambiguity of age and gender complicates moral alignment. She is not an external demon but a distorted reflection of domestic familiarity.

Space is crucial in maintaining this ambiguity. The forest serves as much than just a backdrop for horror; it is also a zone of dependency. Villagers enter it to collect food or supplies, putting themselves in risk. The monster's appearance raises awareness of environmental vulnerability. Children are admonished not only against disobedience, but also against hazardous travel outside communal limits.

This spatial encoding resists flattening into a universal fear template. The narrative's impact stems from its relational density: hunger connects body and land; deception connects trust and vulnerability; and liminality connects home safety and ecological uncertainty. Meaning emerges not from isolated events, but from the interaction of these parts.

Collective Memory and Iterative Voice

Folklore operates through repetition with variation. Each retelling adjusts emphasis based on context, audience, and performance setting. The voice in Nenek Pakande's stories is thus plural and cumulative. No single storyteller owns the story; rather, it is sustained by shared memory. This communal authorship does not abolish accountability, but rather shifts it across generations.

Importantly, oral transmission preserves relational subtlety using performance cues such as intonation, gesture, pauses, and audience response. These embodied aspects help to convey meaning but are challenging to represent through written abstraction. When a tale is written down, some of its relational complexity is eventually lost. Nonetheless, written compilations sometimes include locality markers—place names, distinctive practices, and language textures—that help to anchor the narrative in specific ecological settings.

This iterative voice is in direct contrast to the logic of generative AI. Communal repetition responds to lived conditions, while algorithmic production

adjusts to statistical trends. The former is historically situated, whereas the latter is corpus-driven. While both entail variance, the causes of variation differ significantly.

Monster as Relational Threshold

The figure of Nenek Pakande can thus be understood as a dramaturgical threshold. She marks the boundary at which relational equilibrium falters. Hunger becomes disastrous when it transcends communal regulation; the wilderness becomes dangerous when entered without caution; hospitality becomes hazardous when confidence is misplaced.

Importantly, the narrative rarely simplifies these conflicts to simple dichotomies. The monster may be vanquished, misled, or escaped, but the underlying conditions—scarcity, vulnerability, and environmental risk—continue to exist in communal life. The story serves as relational rehearsal rather than moral closure: a method of visualizing and resolving threats that occur in real life.

From an ecological perspective, this complexity is crucial. The monster is not an externalized evil to be eradicated but a dramatization of imbalance requiring ongoing vigilance. Such narrative structuring resists totalizing resolution and preserves ambiguity as a space of reflection.

Implications for Algorithmic Mediation

Understanding Nenek Pakande as ecological narrative clarifies what is at stake in algorithmic rewriting. If the story is treated merely as horror content, its relational density may be reduced to familiar genre conventions: a terrifying witch, innocent victims, climactic confrontation, moral triumph. While such a retelling may remain interesting, it risks depriving the narrative of its ecological foundation.

The concern, then, is not whether AI can precisely replicate plot elements, but whether it can maintain relationship complexity. Does hunger remain allegorical imbalance, or become cartoonish villainy? Does the forest retain

its ambivalent ecological role, or transform into a generic dark setting? Does communal voice persist, or is the story reframed through individualized heroism?

By first establishing the relational architecture of the folkloric tale, we may develop criteria for judging algorithmic transformation. Ecological dramaturgy does not require textual purity; folklore has always developed. Rather, it inquires if relationship structures—human to land, appetite to scarcity, community to threshold—are maintained, simplified, or altered.

The following part applies this lens to AI-generated rewritings of Nenek Pakande, examining how algorithmic processes alter narrative voice and relational focus. Through comparative research, we can begin to understand how generative systems contribute to the continuing reinterpretation of authorship, originality, and authenticity in the age of artificial intelligence.

Algorithmic Rewriting and Narrative Flattening

While some forms of narrative flattening may appear to result from prompt design—particularly when prompts explicitly specify genre or tone—this study argues that prompts and model tendencies operate relationally rather than independently. Genre-oriented prompts activate statistically dominant narrative patterns embedded within the model's training data, encouraging globally recognizable storytelling conventions over locally situated narrative complexity. Flattening, therefore, emerges not solely from user instruction but from the interaction between prompt guidance and probabilistic alignment.

From Liminal Ecology to Generic Horror

In controlled rewriting prompts, generative AI consistently reoriented the narrative toward recognizable global horror conventions. The forest, previously a site of ambivalent ecological interdependence, was frequently rendered as a dark, ominous setting characterized primarily by danger. Atmospheric descriptors intensified—"shadowy", "twisted trees", "eerie silence"—aligning the narrative with cinematic horror aesthetics.

While such imagery does not inherently distort the story, it shifts relational emphasis. The forest becomes backdrop rather than actor. In folkloric variants, the forest mediates survival and risk simultaneously; villagers depend upon it. In AI-generated versions, dependence is often minimized in favor of suspense. The ecological threshold is simplified into a stage for confrontation.

This transformation illustrates how algorithmic systems gravitate toward statistically dominant genre patterns. As Striphas (2015) argues, algorithmic culture shapes not only what circulates but how forms are structured. When prompted for a "horror retelling", the model mobilizes globally familiar narrative scaffolding, subtly aligning local folklore with transnational templates.

The end result is tone calibration rather than overt erasing. Ambiguousness gives way to clarity, and liminality becomes spectacle. Ecological dramaturgy makes way for genre dramaturgy.

Moral Compression and Narrative Closure

A second pattern observed in AI-generated rewritings is moral compression. Folkloric narratives sometimes contain ambiguity regarding causality and outcome. Nenek Pakande's hunger, while dangerous, can be interpreted as a symptom of imbalance rather than inherent evil. In algorithmic retellings, however, she frequently becomes unequivocally malevolent—a predatory witch whose defeat restores moral order.

This moral sharpening reflects a broader tendency toward narrative closure. AI-generated versions often end with explicit lessons: "Thus, children learned to obey their parents", or "The villagers rejoiced, knowing evil had been defeated". Such endings conform to widely circulating moral tale structures. They provide resolution, satisfying reader expectations shaped by global storytelling conventions.

Yet closure alters relational depth. In traditional variants, survival does not eliminate environmental precarity; the forest remains dangerous, hunger remains possible. In contrast, algorithmic closure means restored equilibrium, reducing complexity to didactic clarity. The shift is subtle but significant: narrative rehearsal becomes moral instruction.

This pattern matches the probabilistic logic of language models. Statistical consistency promotes established story arcs. Unless expressly pushed, ambiguity is generally resolved into coherence. As Crawford (2021) points out, AI systems are taught on corpora that replicate dominant publishing patterns, many of which value clear structure and explicit takeaways.

It is important to distinguish between three overlapping sources of narrative transformation: translation, linguistic dominance, and algorithmic mediation. Some degree of nuance loss may stem from translation itself, as culturally specific expressions and metaphors in the original language do not always have direct equivalents in English. However, beyond this translational shift, AI-generated outputs further exhibit tendencies toward linguistic smoothing and genre alignment, reflecting the dominance of English-language narrative conventions in training corpora. The resulting flattening, therefore, should be understood not as a purely linguistic artifact, but as an effect of interacting layers of mediation—translation, corpus bias, and probabilistic generation.

Individualization of Agency

A third relational shift concerns agency distribution. In communal variants, survival frequently depends on collective vigilance—warnings shared, boundaries respected, experience transmitted. AI-generated rewritings, especially when framed for global audiences, often individualize heroism. A bright youngster or brave villager outwits the monster, turning the story into one of personal triumph.

This reorientation reflects broader narrative conventions in contemporary popular culture, where individual protagonists anchor plot progression. The communal ecology of voice is narrowed to focalized perspective. Such individualization does not eliminate folkloric features, but rather shifts the relational emphasis from collective memory to solo action.

From the standpoint of authorship, this shift is revealing. Folklore distributes narrative authority across generations; algorithmic rewriting may inadvertently reproduce narrative structures that foreground individual protagonists because those patterns dominate training data. Thus, authorship shifts from collaborative relationality to algorithmically managed individualism.

Linguistic Smoothing and Loss of Local Texture

Algorithmic rewriting also tends toward linguistic smoothing. Regional vocabulary, regionally particular metaphors, and structural irregularities typical of oral narration are frequently substituted with universally understandable prose. While such smoothing enhances readability, it may dilute local texture.

This is particularly evident when prompts are given in English. The model's output draws on dominant English-language narrative styles, embedding the story within global literary conventions. Even when key plot elements are preserved, the tonal register shifts toward standardized storytelling norms.

Linguistic smoothing corresponds to the architecture of large language models, which aim for fluency and coherence. Disruptive or unfamiliar structures are statistically less probable and therefore less likely to be generated unless specifically requested. The result is not homogenization by intent but convergence by probability.

Ecological dramaturgy highlights what is lost in this process: relational specificity. Local idioms encode social relations and environmental knowledge. Their replacement with neutral prose alters the narrative's situatedness.

Comparative Excerpts: Original, Translation, and AI-Generated Adaptation

Table 1. Comparative Excerpts: Original, Translation, and AI-Generated Adaptation

Version	Excerpt
Original (Indonesian)	"Demi mempertahankan kesaktiannya, dia harus terus-menerus menyantap daging anak kecil. Ia pun menjelajahi desa demi desa untuk memangsa anak-anak kecil dengan daging yang segar dan lezat".
Translation (English)	"To preserve her supernatural powers, she was forced to continuously consume the flesh of children. She wandered from village to village in search of young victims".
AI--Generated	"Nenek Pakande became feared across the villages as a dark figure who preyed on children to sustain her supernatural existence, spreading terror wherever she wandered".

Source: Compiled by the author from documented Nenek Pakande folklore variants, author's translation, and ChatGPT-generated adaptation.

In the original narrative, Nenek Pakande is not merely represented as a monstrous antagonist but as a figure embedded within ritual and cosmological relations. Her violence is tied to the maintenance of supernatural power, reflecting broader Bugis understandings of sacrifice, spiritual imbalance, and communal vulnerability. Fear is therefore distributed across a collective ecological framework rather than reduced to individual evil.

The English translation retains some of these cultural associations, particularly the connection between supernatural power and sacrificial consumption. However, the translation begins to universalize the story by converting culturally embedded relations into explanatory language more accessible to global readers.

The AI-generated version further standardizes the narrative into globally recognizable horror conventions. Nenek Pakande becomes individualized as a threatening figure defined primarily through terror and predation, while the relational dimensions of ritual obligation, communal ecology, and symbolic ambiguity are significantly reduced. This demonstrates how generative AI systems tend to privilege narrative immediacy and genre familiarity over vernacular complexity.

The following comparative excerpts are presented not as definitive representations of folkloric transformation, but as illustrative examples of how translation and algorithmic rewriting may differently reshape relational density, cultural specificity, and narrative structure.

The AI-generated adaptation was produced using an English-language prompt requesting a concise horror retelling of the Nenek Pakande narrative.

Narrative Flattening as Relational Reconfiguration

Taken together, these patterns—genre alignment, moral compression, individualized agency, and linguistic smoothing—constitute what may be described as narrative flattening. Flattening does not imply superficiality in aesthetic terms; AI-generated texts can be vivid and compelling. Rather, it refers to reduction in relational density.

Relational density is characterized by several crossing tensions, including human-environment interaction, community memory, moral ambiguity, and

geographical liminality. When these tensions are condensed into recognized arcs, complexity is sacrificed in favor of clarity. The monster becomes an archetype rather than a mediator; the forest becomes scenery rather than an ecological actor; and the community becomes a spectator rather than a participant.

Importantly, flattening is not inevitable. Carefully worded prompts can produce more nuanced results. However, the default tendencies of generative systems, as driven by dominant corpora and probabilistic optimization, promote convergence toward globally legible patterns.

This dynamic represents the greater epistemic issue posed by AI. When writing is mediated by statistical abstraction, cultural specificity risks being subsumed beneath generalized narrative schemas. The issue is not technological inability, but structural orientation.

Preparing the Conceptual Shift

The empirical observations above lead directly to the article's central theoretical intervention. If algorithmic rewriting reorganizes relational density, then the challenge is not simply preserving tradition but redefining authorship, originality, and authenticity within hybrid textual ecologies.

Authorship is spread across human prompt designers, model architectures, and training corpora. Originality goes from expressive invention to emergent recombination, guided by curatorial judgment. Authenticity cannot be based merely on textual origins; it must be reframed as relational accountability.

The following section explicitly develops these conceptual implications, linking the comparative findings to broader discussions about writing in the age of AI. This synthesis proposes an ecological model of hybrid authorship that takes into account both technology mediation and cultural distinctiveness.

Rethinking Authorship, Originality, and Authenticity

Authorship: Distributed Computational Agency

Twentieth-century thought shattered the romantic notion of the author as sovereign origin. Barthes (1977) transferred meaning to the reader, while Foucault (1984) reframed authorship as a discursive function. Yet even within these critiques, the text remained anchored in human cultural production. Generative AI introduces a qualitatively different dynamic: textual production without experiential subjectivity.

When a large language model generates a narrative of Nenek Pakande, it does not remember, intend, or interpret. It calculates probabilities across token sequences derived from massive corpora. The resulting text appears intentional, yet intentionality is simulated through statistical coherence. Authorship, in this context, becomes distributed across several layers:

- The human who designs the prompt,
- The engineers who construct the model architecture,
- The dataset curators who assemble training corpora,
- The computational system executing probabilistic inference.

Hayles (2012) describes such configurations as instances of distributed cognition, in which cognitive processes extend beyond individual minds into technical systems. However, distributed cognition does not eliminate asymmetry. The algorithm operates within opaque parameters inaccessible to most users. Agency is shared but unevenly legible.

In folkloric tradition, distributed authorship emerges from communal participation grounded in shared ecology. In algorithmic production, distribution arises from infrastructural abstraction. The distinction matters. Communal authorship retains historical continuity; computational authorship introduces statistical recombination detached from lived relational memory.

Thus, AI challenges authorship not by abolishing it, but by transforming its ontological conditions. The "author" becomes a node within a socio--technical network rather than an embodied narrator. Responsibility must therefore be renegotiated. If the machine produces harmful simplifications or cultural distortions, who answers for them? The prompter? The developer? The institution deploying the system?

The case of Nenek Pakande reveals that authorship in the age of AI cannot be understood solely as origin but must be conceptualized as relational accountability. To author becomes to curate, frame, and ethically position outputs within cultural ecologies.

Originality beyond expressive individualism

Modern literary culture frequently connects originality with innovation based on individual brilliance. But folklore complicates this idea. Repetition and variation are its lifeblood; adaptability is fundamental rather than derivative (Hutcheon, 2006). Originality in oral traditions emerges through contextual inflection rather than unprecedented invention.

Generative AI further destabilizes expressive individualism. Because language models recombine patterns from existing data, their outputs are neither copied nor consciously invented. They are emergent products of probabilistic modeling. This raises a paradox: AI-generated text may be novel in arrangement yet devoid of subjective intention.

The rewriting experiments suggest that AI tends toward statistically dominant narrative forms. When asked to retell Nenek Pakande, the model frequently aligns the story with global horror conventions. Such outputs may appear creatively lively, but their novelty is generally based on surface recombination rather than relational transformations.

As a result, originality transfers from creation to curatorial orchestration. The human contributor's creativity is manifested in generating prompts, selecting outputs, rewriting drafts, and contextualizing the outcome within ethical and cultural constraints. In hybrid writing styles, originality is a result of relational design rather than expressive isolation.

This reconceptualization is consistent with posthuman critiques of individualism. According to Braidotti (2013), subjectivity is always defined by relational networks. In this sense, AI does not eliminate originality; rather, it reveals its relational roots. However, relationality without responsibility increases the risk of homogenization. When statistical regularity dominates, narrative diversity may narrow.

Thus, originality in the age of AI must be redefined as the capacity to sustain relational complexity within technologically mediated production. It is not novelty for its own sake, but attentiveness to ecological and cultural specificity amid probabilistic abstraction.

Authenticity: Relational Transparency

Authenticity is likely the most problematic issue in AI-mediated writing. Can a machine-generated text be considered authentic? The inquiry frequently assumes authenticity as the relationship between text and authorial self. However, folklore complicates this model: the narrative is not anchored by a single ego, and legitimacy is derived on collective recognition rather than personal signature.

AI undermines authenticity by imitating voice without personal experience. A language model can write emotionally compelling writing without experiencing hunger, fear, or ecological precarity. The authenticity dilemma thus concerns relational basis rather than emotional plausibility.

Crawford (2021) reminds us that artificial intelligence systems are founded on extraction infrastructures—data taken, labor concealed, and resources devoured. The "voice" of AI is layered atop these material conditions. When it rewrites folklore, it does so from within globalized linguistic hierarchies. Authenticity is veiled not because the writing is false, but because the conditions of production are unclear.

In the rewriting of Nenek Pakande, authenticity shifts from fidelity to source toward transparency of mediation. A hybrid narrative may be ethically legitimate if it acknowledges technological involvement while remaining focused on relational distinctiveness. Conversely, a seamlessly fluent text that erases its

algorithmic mediation may appear culturally resonant while masking structural simplification.

Ecological dramaturgy redefines authenticity as relational transparency. The question becomes: are the relations that constitute the narrative—human to environment, community to memory, hunger to scarcity—preserved and responsibly represented? Authenticity is less about textual purity than about accountable configuration.

While human authors may also write beyond direct personal experience, their narratives remain grounded in embodied cognition, affective capacity, and socio-cultural positioning. AI-generated writing, by contrast, emerges through probabilistic abstraction rather than situated experience. The distinction therefore lies not between absolute experience and non-experience, but between embodied relational grounding and statistical textual association.

From Decentering to Responsibility

The theoretical arc from Barthes to posthumanism has long emphasised decentering. AI exacerbates this decentering, but it also risks displacing responsibilities. When writing is dispersed across networks, ethical accountability can become elusive.

The study of algorithmic rewriting demonstrates that AI systems tend to standardize, smooth out ambiguity, and match tales with popular genre schemas. These tendencies are not malicious but structural. Therefore, responsibility cannot be abdicated on the grounds that "the machine produced it". Hybrid authorship demands deliberate intervention.

In this framework, the human writer's role evolves. No longer sole originator, the writer becomes relational steward—monitoring how algorithmic processes reshape narrative ecologies. This stewardship encompasses:

- Critical awareness of genre confluence,
- Preserving ecological distinctiveness,
- Resisting unnecessary moral pressure,
- Transparent acknowledgement of mediation.

Such techniques do not restore pre-digital authorship, but rather promote accountable hybridity.

Toward Relational Accountability

AI undermines authorship by redistributing agency, originality by redefining novelty, and authenticity by concealing creation. Yet these challenges also create an opportunity to articulate a relational model of writing.

Relational accountability recognizes that writing in the age of AI is inherently hybrid. The task is not to defend an illusion of isolated human creativity, nor to celebrate algorithmic productivity uncritically. Instead, it is to sustain relational depth amid distributed cognition.

In the context of Nenek Pakande, this means ensuring that rewriting—whether human, algorithmic, or hybrid—remains attentive to ecological dramaturgy. The monster must retain her function as mediator of imbalance rather than devolve into generic villainy. The forest must remain relationally charged rather than aestheticized. The community must remain participant rather than passive backdrop.

Through this viewpoint, authorship becomes an ethical position inside a network of relationships. Originality evolves into curatorial attentiveness to relational complexity. Authenticity evolves into transparency and responsibility in hybrid creation.

The third portion expands on this concept, defining the dimensions of hybrid ecological authorship in response to the algorithmic condition.

Toward Hybrid Ecological Authorship

If generative AI reshapes the ontological conditions of authorship, redefines originality, and destabilizes authenticity, neither a nostalgic return to pre--digital writing nor an uncritical embrace of technology is adequate; instead, a framework that acknowledges technological mediation while sustaining relational accountability is required, which this article conceptualizes as

hybrid ecological authorship. This model recognizes that writing in the AI era is already distributed, with human authors increasingly collaborating—explicitly or implicitly—with algorithmic systems in drafting, revising, translating, and ideation, and thus emphasizes transparency by foregrounding rather than concealing this partnership, situating AI within a relational network of human intention, infrastructural constraints, and cultural memory without granting it moral autonomy. Yet hybridity alone is insufficient, as distributed production can still lead to narrative flattening or cultural uniformity; therefore, the ecological dimension introduces a critical criterion—relational depth—through which writing is evaluated not merely by stylistic fluency but by its capacity to sustain the interdependencies that constitute narrative meaning.

In the case of Nenek Pakande, hybrid ecological authorship entails a series of deliberate practices:

First, it requires attentiveness to ecological specificity. If AI-generated manuscripts tend to follow worldwide horror norms, the human collaborator must intervene to re-center the narrative in its environmental and cultural setting. This may involve restoring ambiguity to the forest, reintroducing communal dynamics, or resisting reductive moral closure.

Second, it calls for preservation of relational multiplicity. Folklore thrives on variation and layered interpretation. Hybrid authorship should avoid collapsing this multiplicity into singularized hero narratives or simplified allegories. Instead, it can use AI-generated variations as exploratory material while retaining critical curatorial judgment.

Third, it demands reflexive positioning. Writers operating within hybrid systems must ask not only what the text says, but how it was produced. Which corpora shape the model's tendencies? Which narrative patterns are statistically amplified? Which voices are marginalized? Such reflexivity is consistent with broader requests for infrastructure awareness in critical AI studies (Crawford, 2021).

Hybrid ecological authorship also transforms pedagogy. In educational settings, AI is frequently viewed as either a cheating device or an efficiency tool. An ecological perspective invites students and scholars to critically engage with AI, investigating how prompts affect narrative construction, genre alignment

arises, and relational complexity is preserved. Writing becomes a negotiation rather than just output generation.

Importantly, this approach does not portray human creativity as necessarily better. Human retellings of folklore can also reduce complexity, impose ideological simplifications, and eliminate ecological nuance. The distinction is between attentive relational configuration and unreflective uniformity, rather than pure human authenticity and manufactured distortion. Both humans and algorithms can influence either outcome.

In this sense, hybrid ecological authorship extends posthumanist insights while reintroducing ethical emphasis. If subjectivity is connected, then accountability must be too. The distributed nature of AI-mediated writing does not absolve accountability; it multiplies its sites.

Practically, hybrid ecological authorship might involve:

- Explicit acknowledgment of AI assistance in creative or academic work,
- Iterative prompting designed to probe rather than finalize narrative structure,
- Comparative reading between human and AI drafts to identify relational shifts,
- Recontextualization of outputs within specific cultural and ecological frameworks.

Through such practices, AI becomes neither autonomous author nor invisible instrument, but participant within a monitored and ethically guided ecology of writing.

In broader cultural terms, hybrid ecological authorship opposes the establishment of narrative monocultures. When algorithmic systems prioritize internationally dominant structures, local traditions risk being assimilated into homogenized templates. Writers can counteract this tendency by consciously curating relational distinctiveness while remaining open to technology mediation.

The reworking of Nenek Pakande becomes more than just a folkloric exercise. It represents a broader shift in how tales spread and change under

computational conditions. Monsters, formerly communicated through embodied communal memory, now travel across digital infrastructures capable of quick duplication. The challenge is not to prohibit such circulation, but rather to ensure that relational density is preserved.

Hybrid ecological authorship provides a path ahead, not defending solitary brilliance or surrendering to computational automation, but rather an ethic of relational stewardship within distributed textual ecologies.

The case of Nenek Pakande indicates that the contact between folklore and generative AI is more than just a technological advancement; it is also a cultural negotiation. Algorithmic systems do not simply replicate narratives; they reorganize narrative ecologies by privileging patterns statistically dominant within training datasets. As a result, localized narrative traditions risk being subtly reshaped through processes of algorithmic normalization.

Recognizing this change enables scholars to go beyond binary discussions about human innovation vs machine automation. Instead, writing in the age of AI might be viewed as a hybrid cultural practice in which narrative meaning arises through the interaction of human interpretation, algorithmic patterning, and cultural memory.

Conclusion

The rise of generative AI compels us to reconceptualize writing as a distributed practice unfolding across socio-technical networks rather than a purely human act, reshaping not only textual production but also the ontological basis of authorship, the epistemology of originality, and the ethics of authenticity. Through the case of Nenek Pakande, this article demonstrates how these abstract shifts become materially visible within vernacular narrative traditions. The comparison between folkloric variants and algorithmic rewritings reveals that AI does not simply reproduce narratives but reorganizes their relational density: ecological ambiguity tends to give way to genre standardization, communal voice shifts toward individualized heroism, moral complexity compresses into didactic closure, and local linguistic textures are smoothed into globally legible prose.

These changes operate subtly through probabilistic alignment with dominant narrative patterns embedded in training data, reflecting a broader condition of algorithmic culture in which statistical regularity risks subsuming narrative particularity. The issue, therefore, is not the production of "false" stories, but the gradual reorientation of story ecologies toward clarity and familiarity at the expense of relational complexity.

In response, this article argues that authorship in the age of AI must be redefined as relational accountability. While poststructuralist and posthumanist thought anticipated the decentering of the author, AI situates this shift within infrastructures that demand renewed ethical attention. Writing becomes part of distributed cognition, yet responsibility must remain explicit through curation, reflexivity, and critical engagement with technological mediation. Originality, in turn, shifts from individual expression to relational orchestration, where creative responsibility lies in sustaining ecological and cultural specificity within hybrid processes. Likewise, authenticity moves beyond essentialist notions of purity toward transparency and the preservation of relational foundations that give narratives depth. The figure of Nenek Pakande exemplifies these stakes: as an ecological mediator, her narrative encodes tensions between scarcity and survival, familiarity and danger—tensions at risk of being flattened through algorithmic rewriting, yet also open to renegotiation within hybrid systems. Rather than casting AI as either threat or solution, this article proposes hybrid ecological authorship as a form of relational stewardship, where writers remain accountable for maintaining narrative complexity within distributed environments. Ultimately, writing in the AI era is defined less by who produces text than by how relationships are configured, sustained, and transformed, positioning the future of narrative within an expanded field of shared responsibility between human and algorithm.

References

Assmann, J. (2011). *Cultural memory and early civilization: Writing, remembrance, and political imagination*. Cambridge University Press.

Barthes, R. (1977/1967). The death of the author. In: *Image, music, text* pp. 142–148). Hill and Wang.

Bauman, R. (1984). *Verbal art as performance*. Waveland Press.

Braidotti, R. (2013). *The posthuman*. Polity Press.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., et al. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. DOI: 10.1016/j.ijinfomgt.2023.102642.

Finnegan, R. (1970). *Oral literature in Africa*. Oxford University Press.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*, 681–694. DOI: 10.1007/s11023-020--09548-1.

Foucault, M. (1984/1969). What is an author? In: P. Rabinow (Ed.), *The Foucault reader* (pp. 101–120). Pantheon Books.

Hayles, N. K. (2012). *How We Think: Digital Media and Contemporary Technogenesis*. University of Chicago Press.

Hayles, N. K. (1999). *How we became posthuman: Virtual bodies in cybernetics, literature, and informatics*. University of Chicago Press.

Hutcheon, L. (2006). *A theory of adaptation*. Routledge.

Kasneci, E., Sessler, K., Küchemann S., Bannert, M., Dementieva, D., Fischer, F., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. DOI: 10.1016/j.lindif.2023.102274.

Striphas, T. (2015). Algorithmic culture. *European Journal of Cultural Studies*, *18*(4–5), 395–412. DOI: 10.1177/1367549415577392.

DOI: 10.2478/doc-2026-0011

This is an open access article licensed under the Creative Commons AttributionNonCommercial-NoDerivatives 4.0 International (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Binh Quoc Duong

Institute of Journalism – Communication, Academy of Journalism and Communication, Viet Nam

jamesduongphoto@gmail.com

ORCID ID: 0009-0002-1396-5672

Ha Thu Le

Institute of Journalism – Communication, Academy of Journalism and Communication, Viet Nam

hale2882@gmail.com

ORCID ID: 0009-0006-0273-2047

Minh Thi Tuyet Nguyen

Institute of Journalism – Communication, Academy of Journalism and Communication, Viet Nam

minhminhnguyen.ajc@gmail.com

ORCID ID: 0009-0001-7056-3103

Visual Politics in the Age of Algorithms: A Four-Dimensional Model of Meaning, Visibility, Affect, and Legitimacy

Article history:

Received **11 December 2025**

Revised **8 January 2026**

Accepted **14 January 2026**

Available online **23 June 2026**

Abstract: This article develops a four-dimensional conceptual framework for analysing visual politics under conditions of algorithmic mediation. It argues that contemporary political images become politically consequential through the interaction of four dimensions that exceed representational analysis alone: constructive meaning-making, regimes of visibility, affective mobilisation, and the visual performance of legitimacy. By synthesising insights from visual culture, political aesthetics, and international relations, the article reframes visual politics as a discursive and infrastructural process shaped by hybrid human-machine assemblages. It further argues, drawing on a contemporary case of platform-circulated synthetic imagery, that algorithmic systems of circulation, ranking, and amplification may reconfigure each dimension, allowing visibility to operate as a provisional source of authority and accelerating affective and legitimacy effects beyond human deliberation. Through analyses of iconic photographs and a case of AI-generated misinformation, the article compares historical configurations of visual politics with a contemporary, platform-mediated episode to clarify

how operative logics shift under algorithmic visibility. The framework offers a transferable analytical vocabulary for examining political imagery in platform--mediated environments and across diverse cultural contexts.

Keywords: visual politics, algorithmic mediation, visibility, affect, legitimacy

Introduction

Images have long been central to political life. From war photography and state ceremonies to protest imagery and humanitarian campaigns, visual forms shape how political events are perceived, remembered, and evaluated. Across visual culture studies, political theory, international relations, and media and communication research, a substantial body of scholarship has demonstrated that images are not merely illustrative supplements to political discourse but active forces in the production of meaning, authority, affect, and collective orientation. Images intervene in what can be seen, felt, recognised, and acted upon in public life, thereby participating directly in the organisation of political experience. As Lehman and Hoyte-West (2021) observe, "novel developments in digital communication … open new avenues of dissemination of knowledge and communication with diversified audiences" (p. 11), a dynamic that likewise reshapes how images circulate across media infrastructures, acquire meaning through dispersed publics, and become politically salient under contemporary conditions of mediated visibility.

Within this expanding field, the concept of visual politics has emerged as a broad analytical lens for examining how power operates through visual practices. Influential contributions have shown that images function as agents of meaning-making (Mitchell, 1994), organise regimes of visibility and authority (Campbell, 2003; Hansen, 2015), mobilise affective responses that shape political judgement (Bleiker, 2009), and establish civic and ethical relations between subjects, spectators, and institutions (Azoulay, 2008). Despite these shared concerns, however, scholarship on visual politics remains conceptually dispersed. Theoretical insights are often developed within distinct disciplinary

traditions and mobilised in parallel, which leaves the field without a consolidated framework capable of capturing how different visual mechanisms interact in the production of political power.

The conditions under which visual politics operates have undergone significant transformation. Contemporary political imagery circulates within media environments increasingly shaped by platform infrastructures, algorithmic curation, and AI-driven image generation. Processes of ranking, recommendation, amplification, and automated moderation now play a central role in determining which images become visible, how they circulate, and what kinds of affective and political responses they generate. Under such conditions, political meaning, authority, and legitimacy are no longer produced solely through human authorship, editorial judgement, or institutional framing, but emerge from hybrid assemblages in which human and computational agents jointly participate. While existing theories of visual politics offer powerful conceptual tools, they have only partially addressed how algorithmic mediation reshapes the production, circulation, and political force of images. These dynamics place questions of authorship, voice, and discursive production at the centre of contemporary visual politics, as images increasingly operate not only as representations but as forms of argument and intervention within algorithmically mediated cultures. Responding to this gap, this article develops a four-dimensional conceptual model of visual politics attentive to such environments, synthesising key strands of visual-political scholarship and re-reading them in light of contemporary media conditions without seeking to replace existing theories.

Drawing on work by Mitchell, Campbell, Hansen, Bleiker, Azoulay, and Hariman and Lucaites, the article conceptualises visual politics as operating across four analytically distinct but interrelated dimensions:

1) the constructive dimension, through which images participate in making political reality;
2) the visibility dimension, through which images organise appearance and concealment;
3) the affective dimension, through which images mobilise emotional response and collective orientation;

4) the legitimacy and power dimension, through which images visualise, perform, and contest political authority.

The model further demonstrates how each of these dimensions is reshaped under conditions of algorithmic mediation, where meaning-making, visibility, affect, and legitimacy are increasingly conditioned by computational systems alongside human actors. Empirically, the article illustrates the framework through a series of iconic case analyses, including *Saigon Execution* (1968), *Falling Man* (2001), and the photograph of Alan Kurdi (2015), supplemented by a contemporary example involving AI-generated misinformation surrounding a fabricated image of an explosion near the Pentagon. These cases demonstrate how visual politics operates across historical moments and media regimes, while also revealing how algorithmic circulation, platform trust signals, and automated amplification introduce new dynamics into the political life of images. Claims about algorithmic mediation in this article are grounded primarily in the contemporary case of platform-circulated synthetic imagery, while the historical photographs establish baseline configurations of visual politics under earlier media regimes.

Literature Review: Key Approaches to Visual Politics in the Age of Algorithms

Images as Agents: Representation, Meaning and Perception

A foundational approach to visual politics begins from the premise that images are not passive reflections of reality but active participants in processes of meaning-making. Mitchell's theorisation of images as cultural agents is central to this view, emphasising that images intervene in perception and interpretation by shaping what becomes thinkable, recognisable, or intelligible within political life (Mitchell, 1994). This emphasis on visual agency reflects a broader epistemological shift toward understanding representation as a constitutive force in political reality. Scholars working in political aesthetics

have extended this insight by examining how shared sensory and perceptual orders structure political possibilities. Blaney's engagement with Rancière, and Bleiker's work on aesthetics and world politics, both foreground the idea that political experience is organised through culturally specific regimes of perception that delimit what can be seen, felt, or acted upon (Blaney, 2023; Bleiker, 2009). From this perspective, visual politics emerges as a domain in which images organise cognitive and symbolic frameworks that anchor political meaning and orient collective understanding. Read through the lens of contemporary media environments, these arguments provide a conceptual foundation for analysing how visual agency is reconfigured when processes of perception and meaning-making are increasingly shaped by algorithmically mediated systems of circulation and visibility.

Visual Authority: Governance, Discourse and Circulation

A second body of research examines how images acquire authority within political discourse and how that authority is governed, contested, and circulated. Campbell's account of cultural governance is pivotal in this regard, arguing that states and institutions do not merely represent political conditions but actively constitute them through visual practices that define what counts as war, threat, or humanitarian crisis (Campbell, 2003). This perspective shifts analytical attention from images as discrete artefacts to the infrastructures and regimes that regulate visibility, including editorial norms, military restrictions, strategic communication, and media logics. Hansen's analysis of international icons extends this line of inquiry by demonstrating how certain images achieve global resonance through repeated circulation, affective uptake, and discursive contestation across political and cultural contexts (Hansen, 2015). Her three--layered model of reception, debate, and appropriation underscores that visual authority is produced relationally, not fixed by any single institutional actor. Read in light of contemporary platform environments, these insights invite further attention to how regimes of visibility are increasingly shaped not only by state and media institutions but also by algorithmic systems of ranking, recommendation, and moderation. In such contexts, visual authority

is reorganised through computational governance, as algorithms participate alongside human actors in determining which images gain prominence, circulate transnationally, or are rendered invisible within public discourse.

Aesthetics and Affect: The Emotional and Sensory Dimensions of the Political

A third strand foregrounds the aesthetic and affective dynamics through which images shape political response. Bleiker's critique of rationalist political epistemologies emphasises that images operate through modes of apprehension that extend beyond deliberation, activating emotional registers that orient political judgment and action (Bleiker, 2009). Affect, in this account, is not reducible to subjective feeling but functions as a structuring force that configures the conditions of political perception. Blaney's work on political aesthetics similarly highlights how sensory experience and affective resonance contribute to the formation of political communities and to the disruption of dominant perceptual orders (Blaney, 2023). Together, these perspectives situate visual politics within broader debates on embodiment and emotion, underscoring that images matter politically because they mobilise feelings that precede, accompany, or exceed rational evaluation. Read in the context of contemporary platform environments, these insights point to the growing significance of algorithmic amplification in shaping affective circulation. In digital media systems organised around engagement and visibility, emotions such as outrage, compassion, and empathy are not merely expressed but are differentially intensified and circulated through algorithmically structured flows, contributing to affective dynamics that influence public attention, collective sentiment, and political orientation.

Photography as a Civic and Ethical Relation

A distinct contribution to visual-politics scholarship is Azoulay's reconceptualisation of photography as a civic encounter. She conceptualises photographs as relational spaces structured by obligations between photographer, photographed subject, and spectator (Azoulay, 2008), shifting

analytical attention away from representation toward civic relations. This notion of a "civil contract" foregrounds the ethical stakes of visibility, particularly in contexts marked by violence, inequality, or dispossession, by emphasising the responsibilities images assign to viewers as civic participants. From this perspective, the political force of photography emerges not simply from what images depict but from the civic relations they enact or foreclose, extending visual politics into questions of citizenship, rights, and transnational ethical claims. Read in light of algorithmically mediated visual cultures, however, Azoulay's framework also raises new questions about responsibility and accountability when images are generated, altered, or circulated by artificial intelligence systems. In cases where no identifiable photographer exists, or where authorship is distributed across datasets, models, and platforms, the ethical relations that underpin the civil contract become less stable and more difficult to locate. This reframing does not invalidate Azoulay's approach but underscores its continued relevance by highlighting how the civic and ethical dimensions of visual politics must be reconsidered under conditions in which human and non-human actors jointly participate in the production and circulation of images.

Algorithmic Mediation and the Transformation of Visual Authorship

Recent debates on AI art and machine vision suggest that visual politics is increasingly shaped not only by representational practices but also by algorithmic mediation, understood as the computational infrastructures through which images are generated, filtered, ranked, and circulated across digital platforms. Instead of functioning as neutral technical tools, algorithms intervene directly in cultural and political processes by structuring visibility, directing circulation, and privileging particular visual forms and narratives. As Zylinska (2020) argues, contemporary AI art needs to be situated within a broader media ecology shaped by platform capitalism, data extraction, and automated decision-making, in which images emerge from complex interactions between human intention, technical systems, and institutional power. This algorithmic turn has significant implications for visual authorship and voice, as AI-generated and algorithmically curated images unsettle long-standing

assumptions about originality, creativity, and individual agency that have historically underpinned photographic and artistic practice. Instead of a singular authorial subject, visual production increasingly takes the form of an assemblage involving datasets, models, interfaces, platforms, and dispersed forms of human labour. Zylinska cautions against framing this transformation through a simplistic opposition between human creators and autonomous machines, calling instead for an approach that foregrounds "entangled human–nonhuman agency" while remaining attentive to the political conditions under which such entanglements operate (Zylinska, 2020, p. 94). From the perspective of visual politics, algorithmic mediation reshapes all four dimensions proposed in this article: it affects constructive meaning-making through generative recombination and algorithmic prioritisation; reorganises regimes of visibility through ranking, moderation, and recommendation; intensifies affective mobilisation through amplification logics tied to engagement; and complicates questions of legitimacy and accountability when visual artefacts are produced and circulated by hybrid human–machine assemblages. This body of scholarship provides a critical bridge between visual theory and contemporary debates on AI, culture, and political communication, underscoring that visual politics in the age of AI is no longer centred on images as discrete artefacts but on the discursive, technical, and institutional processes through which images are made to matter.

Synthesis: Toward an Algorithm-Aware and Discursive Framework

The approaches reviewed in above sections collectively demonstrate that visual politics has developed as a multifaceted field concerned with how images participate in the production of political meaning, authority, affect, and ethical relations. Although these strands of scholarship emerge from different disciplinary traditions (visual culture, political aesthetics, international relations, and media studies), they converge on an understanding of images as active forces embedded in broader cultural and institutional processes. Contemporary debates increasingly emphasise that these processes now unfold within environments profoundly shaped by algorithmic mediation.

Across the literature, several interrelated analytical mechanisms become apparent. Research on images as agents of meaning-making underscores that political interpretation is never a matter of representation alone but is structured through culturally organised regimes of perception. In algorithmically mediated contexts, this insight extends to the ways in which filtering, prioritisation, and circulation condition what becomes intelligible or salient within public discourse. Meaning, in this sense, is produced not only through symbolic content but also through the infrastructural arrangements that govern visual exposure.

Studies of visual authority and governance further demonstrate that visibility operates as a politically regulated resource. While earlier work focused primarily on states, media institutions, and editorial practices, more recent scholarship highlights the growing influence of platform logics and computational systems. Algorithmic ranking, recommendation, and moderation increasingly function as regimes of visibility that shape transnational circulation and redistribute authority over which images gain prominence or remain marginal. Authority is thus neither fully centralised nor dissolved but reconfigured through hybrid forms of human and computational governance.

Aesthetic and affect-oriented approaches, in turn, emphasise that images exert political force by mobilising emotions that orient judgment and collective response. In digital media environments organised around engagement and virality, affective dynamics are not merely intensified but patterned through algorithmic amplification. Emotions such as outrage, empathy, or compassion circulate unevenly, contributing to the formation of affective publics whose attention and orientation are shaped by platform-driven logics of visibility. Affect thereby functions as a crucial link between visual experience and political mobilisation under contemporary conditions.

Ethical and civic perspectives, particularly Azoulay's conception of photography as a civil contract, draw attention to the responsibilities and relational obligations embedded in acts of seeing. Algorithmically mediated and AI-generated imagery complicates these relations by dispersing authorship and obscuring lines of accountability. When images are produced or altered by hybrid human-machine assemblages, questions of responsibility, legitimacy, and ethical obligation become more difficult to locate, yet more

politically consequential. Instead of rendering civic approaches obsolete, these developments underscore the need to reconsider how ethical relations are constituted under conditions of distributed agency.

These approaches indicate that visual politics can no longer be adequately analysed through images as isolated artefacts or through singular sites of production and interpretation. Visual power increasingly emerges from discursive processes that unfold within complex assemblages of human actors, technological infrastructures, and institutional arrangements. Algorithmic mediation does not replace established theories of visual politics but reconfigures the conditions under which they operate, reshaping how images are produced, circulated, encountered, and made politically consequential.

In this context, visual politics is best understood not as a stable object of analysis but as a dynamic field of practices in which agency is distributed across human and non-human actors. Algorithms do not simply transmit or distort visual meaning; they participate in structuring regimes of perception, amplifying affective responses, governing visibility, and complicating questions of authorship and accountability. What is at stake, therefore, is not the emergence of a wholly new form of visual politics, but a transformation in its operative logic: from image--centred analysis to an examination of the discursive, technical, and institutional processes through which visual meaning, authority, and responsibility are continuously negotiated in algorithmically mediated environments.

This synthesis provides the conceptual groundwork for the following section, which articulates a definition of visual politics attentive to these transformations and introduces a multidimensional analytical framework for examining the interconnections between meaning, visibility, affect, and legitimacy under contemporary conditions.

Defining Visual Politics in Algorithmically Mediated Environments

Understanding what scholars mean by visual politics requires navigating a field that spans multiple disciplines yet remains conceptually dispersed.

Research in visual culture, political theory, international relations, and media and communication studies has consistently demonstrated that images play a central role in shaping how publics perceive, interpret, and engage with political life. Despite this sustained attention, however, the term *visual politics* has often been employed descriptively rather than defined analytically, leaving its conceptual centre of gravity under-specified. Building on the theoretical synthesis developed in the previous section, this section articulates a definition of visual politics attentive to the transformations introduced by algorithmic mediation while preserving the core insights of established visual theory.

A foundational point of departure lies in the rejection of purely representational accounts of images. As Mitchell argues, images are never neutral surfaces for the transmission of meaning but participate in a "politics of representation" that shapes what can be thought, said, and recognised within a given cultural moment (Mitchell, 1994). This understanding is reinforced by Rancière's theorisation of the "distribution of the sensible", which conceptualises politics as a continual reconfiguration of what can be seen, heard, and rendered intelligible in public life. From this perspective, visual politics does not concern images as static objects but as practices that intervene in perceptual orders and delimit the conditions of appearance. Images matter politically not only because of what they depict, but because of how they organise attention, structure interpretive possibilities, and regulate visibility within shared cultural spaces.

Work in international relations and political communication further extends this argument by demonstrating that visual representation is constitutive of political reality. Campbell shows that images help produce and stabilise categories such as war, threat, and humanitarian crisis, functioning as mechanisms through which political authority is imagined, legitimised, and contested (Campbell, 2003). Hansen's analysis of international icons similarly reveals that the political significance of images emerges through processes of circulation, emotional resonance, and discursive uptake across national and cultural contexts (Hansen, 2015). These accounts underscore that visual politics operates through practices of governance and circulation, through which images are embedded in broader regimes of visibility and authority.

Aesthetic and affect-oriented approaches highlight that the political force of images cannot be reduced to cognitive interpretation or discursive framing. As Bleiker argues, images mobilise affective responses that orient judgment and action through sensory and emotional registers that precede or exceed rational deliberation (Bleiker, 2009). Visual politics thus unfolds partly through embodied experience, shaping political dispositions by aligning emotions such as fear, outrage, empathy, or solidarity. In contemporary media environments, where images circulate rapidly across digital platforms, these affective dynamics are increasingly structured by algorithmic systems that amplify, prioritise, or suppress particular emotional cues, further complicating the relationship between visual experience and political engagement.

Azoulay's reconceptualisation of photography as a civic encounter adds an ethical and relational dimension to this field. By framing photography as a "civil contract" between photographer, photographed subject, and spectator, Azoulay foregrounds the responsibilities and obligations embedded in acts of seeing (Azoulay, 2008). Visual politics, from this standpoint, is inseparable from questions of citizenship, rights, and recognition, as images position viewers within civic relations that entail ethical claims. Under conditions of algorithmic mediation and AI-generated imagery, however, these relations are increasingly destabilised. When authorship is distributed across datasets, models, interfaces, and platforms, responsibility and accountability become more difficult to locate, raising new questions about agency and legitimacy in visual communication.

These perspectives converge on several shared commitments across the field:

1) images are treated as agents in the production of political reality, not merely as neutral carriers of information;
2) visibility is understood as politically structured, with images regulating who appears in the public sphere and under what interpretive conditions;
3) affect is recognised as a central mechanism through which images shape political orientation and collective response;

4) visual practices establish relational and ethical claims that bind image-makers, depicted subjects, and publics within fields of responsibility and recognition; and
5) images contribute to the formation and reconfiguration of political communities through shared memory, symbolic identification, and affective mobilisation.

Under contemporary conditions of algorithmic mediation, these commitments remain foundational but must be reconsidered in light of distributed agency and computational governance. Building on these converging insights, visual politics may be defined as the study of how visual practices and representations produce, distribute, and contest political power by shaping perception, visibility, affect, and relations of agency and responsibility within culturally, institutionally, and algorithmically mediated environments. This definition foregrounds images not as isolated artefacts but as participants in discursive and technical processes through which political meaning, authority, and legitimacy are continuously negotiated. By conceptualising visual politics as a multidimensional and processual field, this approach provides the foundation for the analytical framework developed in the following section.

A Four-Dimensional Conceptual Model of Visual Politics

Constructive Dimension: How Images Participate in Making Political Reality

Images function as formative elements in the production of political meaning, shaping political reality beyond any role as neutral reflections of external events. Across the theoretical traditions reviewed above, scholars emphasise that visual materials intervene in how publics understand events, assign moral significance, and construct collective memory. Mitchell's account of images as culturally operative agents highlights their capacity to organise perception and delimit

interpretive possibilities, shaping what becomes knowable, recognisable, and politically salient (Mitchell, 1994). From this perspective, images participate actively in the construction of political reality by providing symbolic frameworks through which events are interpreted and evaluated.

This constructive role is further developed in Campbell's analysis of visual representation in international politics, which demonstrates that images actively constitute phenomena such as war, security, or humanitarian crisis, instead of simply depicting them (Campbell, 2003). Images do not simply illustrate political conditions; they shape the conceptual and moral vocabularies through which those conditions become intelligible. Hansen's notion of international icons extends this insight by demonstrating how certain images acquire political force through circulation, resonance, and discursive uptake across different cultural and geopolitical contexts (Hansen, 2015). Through repeated circulation and affective engagement, such images offer condensed interpretive templates that stabilise particular understandings of political events while marginalising alternative readings. The constructive dimension therefore refers to the capacity of images to anchor narratives, structure moral frameworks, and supply the symbolic resources through which political realities take form. This work is simultaneously epistemic and affective: images shape both what is known and how it is felt, operating across cognition, memory, and shared cultural imagination. Political meaning emerges not solely from visual content but from the ways images are embedded within discursive practices of interpretation, circulation, and remembrance.

Under contemporary conditions of algorithmic mediation, this constructive function is increasingly shaped by computational systems that participate in the generation, recombination, and prioritisation of visual forms. Meaning-making is no longer fully attributable to human authorship but emerges from assemblages involving datasets, models, platforms, and interpretive communities. In such environments, the constructive dimension of visual politics extends beyond representational framing to encompass algorithmically mediated processes that influence which images are produced, how they circulate, and which interpretations gain prominence within public discourse.

Visibility Dimension: The Politics of Appearance and Concealment

The second dimension concerns the ways images organise visibility, determining what enters public perception and what remains obscured. Politics, as Rancière suggests, is inseparable from the distribution of the sensible: the normative arrangements that govern who may appear, which experiences become recognisable, and which forms of life become intelligible as part of public life. Viewed through this lens, images function not merely as representations of reality but as mechanisms that regulate access to visibility, shaping the perceptual boundaries within which political subjects and events can be recognised.

Azoulay's emphasis on absence as a constitutive feature of photography deepens this perspective by drawing attention to what is excluded from the frame (Azoulay, 2008). What remains unseen, whether through structural marginalisation, censorship, or editorial decision-making, can be as politically consequential as what is made visible. Absence here should be understood not as a lack, but as a condition actively produced through visual practices. Campbell's analysis of visual regimes further demonstrates how states, media institutions, and other authoritative actors manage these boundaries of appearance, authorising certain images of suffering, threat, or authority for public circulation while suppressing others (Campbell, 2003).

The visibility dimension thus highlights that images exercise political force by organising perceptual hierarchies. They shape recognition, determine whose experiences count as publicly intelligible, and delineate the boundaries of collective attention. Visibility is therefore not merely a descriptive condition but a distributive practice through which political subjectivity is affirmed or denied. Lorés (2021) reminds us that "visibility can only be properly understood if verbal markers are analysed in combination with other modes with which they connect (i.e. visual)" (p. 53), underscoring that visibility is always a multimodal and discursively mediated phenomenon.

In contemporary platform-mediated environments, regimes of visibility are increasingly shaped by computational systems that rank, recommend, and moderate visual content across digital platforms. Visibility becomes a dynamically governed resource, structured not only by institutional decision-making but

also by platform logics that prioritise engagement, circulation, and attention. Algorithms do not simply transmit images; they participate in organising what becomes visible, marginal, or invisible at scale within platform-mediated publics. In such environments, the politics of appearance and concealment is reconfigured through hybrid regimes in which human and computational forms of governance jointly shape the visual field of political life.

Affective Dimension: Emotional Economies of Visual Communication

A third dimension centres on the affective charge of images, their capacity to evoke and circulate emotions that become integral to political engagement. Unlike verbal discourse, which tends to operate through rational deliberation, images frequently intervene through immediacy and intensity. As Bleiker argues, aesthetic experience opens onto aspects of political life that remain inaccessible to rationalist frameworks, reaching publics through sensory and emotional registers that precede reflective judgment (Bleiker, 2009). From this perspective, affect is not a secondary response to visual content but a constitutive mechanism through which political meaning is apprehended and oriented.

Hansen's work on international icons further underscores that images acquire symbolic and political force when they activate strong emotional identification (Hansen, 2015). Their resonance is not reducible to representational content but depends on affective intensity and circulation. Communication research similarly suggests that publics often rely on affective heuristics when engaging with political imagery, particularly in contexts characterised by uncertainty, distance, or low levels of prior information. The affective dimension thus refers to the emotional infrastructures within which images circulate, resonate, and acquire political significance.

Within digital media environments organised around engagement and virality, affective dynamics are increasingly patterned through algorithmic amplification. Images that provoke outrage, compassion, fear, or empathy are differentially prioritised and circulated according to platform logics that reward emotional intensity and sustained attention. In such contexts, affect does not merely accompany visual communication but becomes a key driver of

visibility, circulation, and political mobilisation. The affective dimension of visual politics therefore operates through feedback loops in which emotional response, algorithmic prioritisation, and public attention mutually reinforce one another, shaping how political images are encountered, interpreted, and acted upon.

Legitimacy and Power Dimension: Visualising Authority and Political Order

The fourth dimension highlights how images participate in the articulation and contestation of political authority. Visual representations do not merely depict power; they contribute to its performance, stabilisation, and recognition within public life. Contemporary politics is increasingly mediated through style, symbolism, and visual display. The self-presentation of political actors (through posture, gesture, setting, and symbolic alignment) functions as a form of visual performance through which credibility, authority, and legitimacy are constructed and communicated.

Hariman and Lucaites (2016) emphasise that photographs stage civic identity and invite viewers into shared interpretive spaces where legitimacy is either affirmed or challenged. Images operate as sites of civic spectatorship, positioning publics in relation to authority through visual cues that signal trust, leadership, or moral standing. Campbell's notion of pictorial authority similarly captures how visual regimes naturalise state narratives, choreograph scenes of unity and purpose, and render particular political orders as self-evident or necessary (Campbell, 2003). In these accounts, legitimacy emerges not from coercion alone but from the successful visual performance of authority within culturally recognisable forms. The legitimacy and power dimension therefore concern the aesthetic and symbolic labour through which political authority becomes visible, persuasive, and publicly recognisable. Images help stabilise political orders by aligning actors, symbols, and emotions within coherent visual narratives, yet they also expose points of fragility when performances fail or when alternative visual representations contest dominant claims to authority. Authority, in this sense, is always provisional and dependent on ongoing processes of visual recognition.

In contemporary media environments marked by AI-generated imagery, synthetic media, and algorithmically intensified circulation, the visual performance of authority becomes increasingly unstable. The proliferation of manipulated, staged, or automated images complicates established assumptions about authenticity, authorship, and accountability. Political legitimacy can no longer rely solely on institutional credibility or human authorship, but must instead be negotiated within hybrid human–machine assemblages in contemporary visual cultures, where authority is performed, circulated, and evaluated under conditions of heightened visual uncertainty. In such contexts, the legitimacy of political power is not only visualised but continually tested through shifting dynamics of trust, recognition, and technological mediation.

Interactions Across the Four Dimensions

Although each of the four dimensions (constructive, visibility, affective, and legitimacy) captures a distinct analytical mechanism, their political effects are realised only through their interaction, as visual politics rarely operates through any single dimension in isolation. Images acquire political significance through the interplay of these dimensions, as processes of meaning-making, visibility, emotion, and authority reinforce or destabilise one another within specific cultural, communicative, and institutional settings. The political force of images thus emerges not from isolated visual effects but from the cumulative and relational dynamics through which images are produced, circulated, encountered, and evaluated.

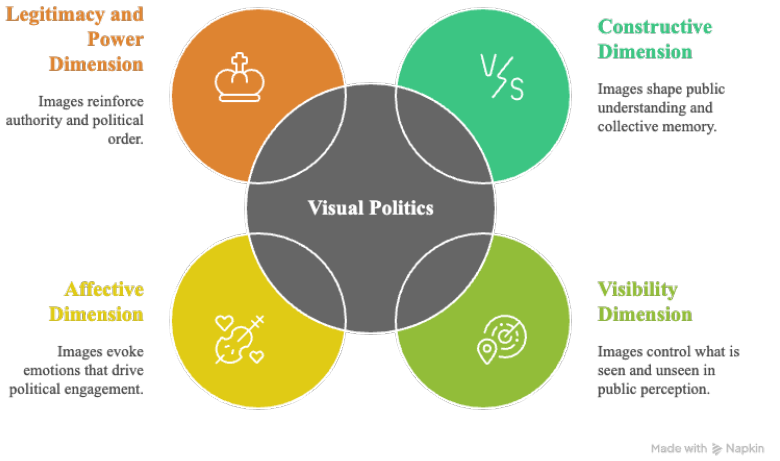
An image that shapes political reality (the constructive dimension) typically does so because it also structures what is rendered visible or invisible (the visibility dimension). Crucially, the interpretive power of a photograph depends not only on what it shows but also on what it withholds, and such selective framing determines whether particular actors, grievances, or narratives enter public consciousness. These regimes of visibility, in turn, condition the affective responses images invite. Strong emotional reactions, such as shock, empathy, or indignation, often arise from the tension between

appearance and exclusion, illustrating how affective dynamics build upon and intensify the politics of visibility.

The emotional force of images further feeds into questions of legitimacy and authority. Images that mobilise collective feeling can consolidate political orders by affirming shared values and stabilising claims to authority, or conversely, expose ruptures in public trust by visualising vulnerability, failure, or abuse. The legitimacy dimension thus emerges from the interaction of affective resonance, symbolic construction, and authoritative visibility. Political power becomes intelligible and persuasive not through any single visual operation, but through the interdependent alignment of meaning, emotion, and recognition.

In contemporary media environments, these interactions increasingly unfold within algorithmically mediated systems of circulation and governance. None of the four dimensions operates outside such environments. Processes of meaning-making, for instance, are shaped by algorithmic recombination and prioritisation; visibility is governed through ranking, recommendation, and moderation; affect is amplified through engagement-driven feedback loops; and legitimacy is negotiated under conditions of distributed authorship and heightened uncertainty. The figure below conceptualises these dynamics not as a static model, but as a relational framework through which visual politics operates across hybrid human–machine assemblages.

Figure 1. A Four-Dimensional Conceptual Model of Visual Politics in Algorithmically Mediated Environments



Source: Author's own elaboration based on the conceptual framework developed in this article.

Understanding visual politics therefore requires an integrated analytical approach attentive to how these dimensions interact under contemporary conditions. A multidimensional framework enables researchers to trace how images move across cultural, technological, and political contexts, activating different dimensions at different moments. It also clarifies why the same image may stabilise one political configuration while destabilising another, depending on how processes of meaning-making, visibility, affect, and legitimacy are articulated within algorithmically mediated environments. Recognising these interactions provides a more comprehensive account of the political labour images perform and establishes the conceptual foundation for the case analyses that follow.

Case Analyses: Demonstrating the Framework

Saigon Execution (1968): Constructing a Political Reality Through Shock

Figure 2. Eddie Adams, *Saigon Execution* (1968)

Source: Associated Press (AP Photo/Eddie Adams).

The photograph commonly known as *Saigon Execution*, taken by Eddie Adams on 1 February 1968, has become one of the most widely circulated visual documents of the Vietnam War. Instead of functioning merely as a record of a single moment of violence, the image came to symbolise broader tensions surrounding the moral and political legitimacy of the conflict. Its significance lies not only in the event it depicts, but in the interpretive frameworks, cultural narratives, and public emotions that crystallised around it over time.

Constructive dimension

The image rapidly became a reference point in emerging critiques of the U.S. war effort. Scholars and commentators have frequently interpreted it as contributing to shifts in how American publics imagined the conflict, condensing complex geopolitical dynamics into a single, morally unsettling scene. In this sense, the photograph exemplifies Mitchell's and Campbell's claims that images participate in shaping political realities by reconfiguring the symbolic and affective conditions under which events become intelligible (Mitchell, 1994; Campbell, 2003). Its iconicity does not derive from informational value but from its capacity to anchor powerful narratives about brutality, ambiguity, and the ethical stakes of war.

Visibility dimension

The visual structure of the image foregrounds certain actors while leaving crucial contextual details outside the frame. This selective visibility aligns with Rancière's notion of the distribution of the sensible, in which political meaning arises from the organisation of appearance and disappearance. Azoulay's emphasis on the politics of the frame further clarifies how the absence of background information, including the circumstances leading to the execution and the identities of those involved, shaped the photograph's long afterlife, enabling it to circulate as a symbolic condensation rather than a documentary explanation (Azoulay, 2008).

Affective dimension

The emotional jolt produced by *Saigon Execution* played a central role in its cultural circulation. Scholars such as Hariman and Lucaites (2016) have shown that iconic images generate shared affective environments in which dispersed publics momentarily converge. The shock, disorientation, and moral discomfort triggered by the photograph allowed it to resonate across ideological boundaries, becoming a touchstone in public debates about violence, responsibility, and the human cost of war.

Legitimacy and authority

The image also became entangled in negotiations over political authority. By making visible an act of violence committed by an allied general, the photograph unsettled dominant narratives that framed the conflict in terms of strategic necessity or moral clarity. The photograph did not directly cause political change; instead, it provided a visual vocabulary through which scepticism, anger, and ethical questioning could be articulated. Its continued use in textbooks, museum exhibitions, and documentary films demonstrates how the image has shaped cultural memory, becoming part of a long-standing discourse on war, representation, and accountability.

Saigon Execution illustrates a regime of visual politics in which political meaning, affect, and legitimacy were produced through editorial circulation, print journalism, and historically situated interpretation. These conditions establish an important baseline for understanding how visual politics operated prior to the emergence of platform-driven visibility and algorithmic amplification examined later in this article.

Falling Man (2001): The Politics of the Unshown

Figure 3. Richard Drew, *Falling Man* (2001)

Source: Associated Press (AP photo/Richard Drew).

Richard Drew's *Falling Man* occupies a distinctive place in the visual history of 9/11. Briefly circulated on the day of the attacks before being withdrawn by many U.S. news organisations, the image became emblematic not of what was widely seen, but of what was rendered unseeable. Its trajectory illuminates the politics of visibility, taboo, and national mourning that shaped how the attacks were publicly remembered and emotionally processed.

Constructive dimension

Even in its partial circulation, the image introduced a representational possibility that stood in tension with emergent narratives of heroism, resilience, and

collective unity. Scholars have noted that *Falling Man* became an inadvertent counterpoint to official discourses, quietly haunting cultural imagination despite its absence from mainstream news coverage. Its later reappearances in literature, art, and documentary film suggest that the image continued to shape interpretive frameworks long after its initial suppression, functioning as a latent visual reference for experiences of vulnerability and loss that could not be fully accommodated within dominant narratives.

Visibility dimension

The withdrawal of the photograph exemplifies the cultural management of visibility in moments of national crisis. Decisions not to publish were framed as gestures of respect and sensitivity, yet they simultaneously stabilised a discourse of mourning in which certain forms of suffering, particularly those associated with desperation, fear, or ambiguity, were rendered unacceptable for public display. Rancière's notion of the distribution of the sensible helps explain how such omissions contribute to the formation of collective memory by delimiting the emotional and symbolic boundaries of what may be seen and acknowledged. Azoulay's concept of the "missing people of photography" further clarifies how those excluded from representation nonetheless remain central to the political meaning of the event, haunting public memory through their very absence (Azoulay, 2008).

Affective dimension

The emotional intensity of *Falling Man* was a key factor in its disappearance from public circulation. Its stark portrayal of individual vulnerability disrupted predominant affective registers of heroism, resolve, and national strength that quickly came to dominate visual narratives of 9/11. As Bleiker and Hansen suggest, images that generate affective dissonance are often perceived as incompatible with stabilising political narratives, prompting their removal from public space (Bleiker, 2009; Hansen, 2015). In this sense, *Falling Man* illustrates not only how

affect circulates but how emotional responses are actively regulated, managed, and selectively excluded within visual regimes of crisis.

Legitimacy and authority

The suppression of *Falling Man* also played a role in shaping the legitimacy of the political responses that followed the attacks. By narrowing the visual field to images that affirmed unity, resilience, and moral clarity, media institutions contributed to the stabilisation of narratives that underpinned emerging security agendas and military interventions. Yet the image's later recirculation, particularly in cultural and artistic contexts, reveals enduring tensions over how authority, vulnerability, and national identity are visually articulated. Its afterlife underscores the long-term cultural negotiation of grief, trauma, and the politics of remembrance.

Falling Man exemplifies a historically specific regime of visual politics in which visibility and absence were governed primarily through editorial judgement, professional norms, and culturally embedded expectations of propriety and sensitivity. This case demonstrates how political meaning and legitimacy can be shaped as much by what is withheld from view as by what is displayed, providing an important historical reference point for understanding later transformations in the governance of visibility addressed elsewhere in this article.

Alan Kurdi (2015): Global Affective Shock and the Recalibration of Responsibility

Figure 4. Nilüfer Demir, *Alan Kurdi* (2015)

Source: Associated Press/DHA (AP Photo/ Nilüfer Demir).

The photograph of Alan Kurdi, taken by Nilüfer Demir in 2015, rapidly became one of the most widely recognised images of the Syrian refugee crisis. Its impact illustrates how visual politics operates across transnational media environments in which meaning, emotion, and responsibility are negotiated among diverse publics. Unlike images whose circulation was confined primarily to national media systems, the photograph of Alan Kurdi travelled swiftly across borders, platforms, and interpretive communities, acquiring political significance through its global visibility and emotional resonance.

Constructive dimension

The image intervened in dominant European framings of migration, which had largely revolved around security, regulation, and demographic anxiety. Scholars and commentators have interpreted the photograph as reshaping the discursive landscape by shifting attention from abstract categories such as "flows" or "burdens" to a singular, embodied human tragedy. In Campbell's terms, the image offered an alternative way of imagining the crisis, reconfiguring narratives and moral vocabularies through which "the refugee" became publicly understood (Campbell, 2003). Political meaning emerged less from policy argumentation than from the condensation of complex geopolitical conditions into an emotionally charged visual encounter.

Visibility dimension

The photograph made visible forms of suffering that had long remained structurally marginal within European public discourse. Thousands of people had died in the Mediterranean prior to Alan Kurdi, yet such deaths rarely entered the visual field of European publics. By foregrounding a child's body in a peaceful, almost serene setting, the image disrupted established tropes of disorder, threat, or illegality commonly associated with migration. In doing so, it reconfigured the distribution of the sensible by bringing vulnerable bodies into public visibility and reframing the moral terms under which the refugee crisis was recognised.

Affective dimension

The global affective response to the photograph was immediate and intense. Widespread sharing on social media, front-page newspaper coverage, and a surge in charitable initiatives followed its circulation. The image mobilised a collective emotional shock that temporarily reshaped public conversation through affective immediacy, with limited engagement in sustained deliberation or policy analysis. As Hansen's and Bleiker's accounts of affective activation suggest, the photograph condensed complex political and humanitarian issues into

a singular emotional encounter, aligning dispersed publics through shared feelings of grief, empathy, and moral urgency (Hansen, 2015; Bleiker, 2009).

Legitimacy and political authority

The image was frequently invoked in political commentary and public debate and has been linked to short-term shifts in policy discourse and refugee commitments, particularly in countries such as Germany and Canada. Hariman and Lucaites' concept of civic spectatorship is useful here, as viewers were positioned not as distant observers but as ethical participants addressed by the scene. At the same time, responses varied across cultural and national contexts. In parts of Western Europe, the image catalysed compassion and calls for responsibility; in parts of Eastern Europe, it provoked anxieties about national identity; and in other regions it circulated within narratives of regional suffering and geopolitical abandonment. These divergent receptions underscore how the political meaning of images is negotiated across culturally specific interpretive frameworks.

The photograph of Alan Kurdi exemplifies a transitional moment in visual politics. Its circulation relied on both traditional journalistic infrastructures and emerging social media platforms, combining editorial selection with rapid peer-to-peer dissemination. The image demonstrates how affective mobilisation and civic address operated within a media environment that was increasingly networked but not yet fully governed by algorithmic amplification. As such, it provides an important historical bridge between earlier regimes of editorial visibility and the algorithmically mediated visual politics examined in the following section.

Pentagon Explosion AI Image (2023): Algorithmic Visibility and Provisional Legitimacy

Figure 4. AI-generated image falsely depicting an explosion near the Pentagon (2023)

Source: Social media circulation; analysed with reference to Clayton (2023).

The rapid circulation of an AI-generated image falsely depicting an explosion near the Pentagon in May 2023 provides a revealing empirical test of the proposed framework. Although the image lacked any documentary referent and was rapidly debunked in subsequent reporting, its dissemination through multiple verified social media accounts generated immediate political and economic effects, including automated news alerts, international media pickup, and a short-lived dip in U.S. stock markets (Clayton, 2023). Under algorithmically mediated conditions, political meaning and authority may emerge independently of representation, authorship, or institutional validation, as this episode makes visible. Visibility itself can function as a provisional source of legitimacy, as algorithmic amplification and platform trust signals enable images to produce real-world consequences before processes of interpretation, verification, or ethical judgment can intervene. The case therefore exposes the fragility of established regimes of visual authority in environments where circulation, credibility, and response are increasingly organised through computational infrastructures.

Constructive dimension

In this case, political meaning was activated without any stable referential grounding. In contrast to the historical images examined earlier in this article, whose political force emerged through their embedding in specific events and discursive contexts, the Pentagon image produced meaning procedurally. The image generated the assumption that an event had occurred. Meaning arose from patterns of circulation, platform cues, and inferential logics, with symbolic framing operating under severe time constraints and often lagging behind circulation. The constructive power of the image thus resided in algorithmically mediated inference: within platformed publics, if an image is rapidly circulating across trusted platforms, it is provisionally treated as meaningful and potentially true. This form of constructive meaning-making illustrates a shift from symbolic condensation to procedural activation within contemporary visual politics.

Visibility dimension

The visibility of the Pentagon image emerged through platform-driven amplification, bypassing editorial selection and institutional validation. Verification badges, real-time cross-platform circulation, and repetition within automated aggregation systems combined to produce a regime of visibility structured by platform infrastructures alongside institutional signals of authority. The image achieved large-scale exposure prior to any sustained verification, allowing visibility itself to operate as a provisional source of legitimacy. Authority in this configuration emerged from speed, reach, and platform-mediated trust cues, as algorithmic circulation elevated the image's political relevance ahead of journalistic or governmental stabilisation. Here, circulation functions as a practical condition through which political relevance is initially recognised.

Affective dimension

The affective force of the image was generated less through emotional identification than through temporal compression and uncertainty. The sudden

appearance of a visually plausible image depicting a major security incident produced a momentary shock rooted in ambiguity and urgency rather than empathy. Unlike the affective mobilisation associated with images such as the photograph of Alan Kurdi, which relied on compassion and moral appeal, the Pentagon image activated affect through acceleration. The rapid succession of alerts, reposts, and media pickups compressed reaction time, producing a short-lived affective environment characterised by anxiety and immediacy. In algorithmically mediated contexts, affect is increasingly structured by speed and simultaneity, as automated circulation intensifies emotional response before reflective judgment becomes possible.

Legitimacy and power

The most consequential implication of the case concerns a reconfiguration of legitimacy under algorithmic conditions. Political and economic effects unfolded before verification, institutional response, or ethical evaluation could stabilise interpretation, indicating a form of legitimacy that was contingent, reversible, and sustained through algorithmically produced visibility. The episode illustrates how political power can be exercised through machine-mediated visual performance, in which infrastructural signals generate authoritative effects at a pace that exceeds deliberative and regulatory processes.

Reconfiguring the Four Dimensions under Algorithmic Conditions

The Pentagon image illustrates how the operative logic of visual politics is reconfigured under algorithmic conditions, instead of indicating the emergence of an entirely new form. The core mechanisms identified throughout this article, constructive meaning-making, regimes of visibility, affective mobilisation, and the performance of legitimacy, remain analytically recognisable in this case. What changes is their sequencing, speed, and source of authority. Historical cases illustrate how visual politics operated through symbolic condensation, editorial governance, and culturally mediated affect. The contemporary case shows how these same mechanisms are reconfigured within algorithmic

regimes in which visibility is produced first, affect follows immediately, and legitimacy is provisionally assigned before interpretation or verification can occur. This reordering clarifies how visual politics unfolds within hybrid human and machine assemblages, where computational infrastructures participate directly in shaping political meaning, authority, and response.

Discussion

Conceptual Contributions

The four-dimensional approach articulated in this article brings together theoretical strands that have often been treated in isolation within the study of visual politics. The model shows how their insights address distinct yet interrelated mechanisms through which images acquire political force. This perspective suggests that visual politics is best understood through a constellation of discursive operations (constructing meaning, organising visibility, mobilising affect, and performing legitimacy) that unfold simultaneously across cultural, media, and institutional environments. The framework also clarifies how images function as forms of argument and discursive intervention, particularly under conditions of algorithmic amplification. By foregrounding these operations, the article shifts analytical attention from images as discrete objects to the practices and conditions that enable them to circulate, resonate, and matter politically. This move helps clarify a field in which terminologies and analytical emphases have developed unevenly, and it provides a shared vocabulary for examining how visual meaning, authority, and emotional response are produced across different domains of political communication.

The framework also extends existing theories of visual politics by situating these operations within contemporary environments increasingly shaped by algorithmic mediation. Meaning-making, visibility, affective resonance, and claims to legitimacy are no longer produced solely through human authorship or institutional control, but emerge from hybrid assemblages involving platforms, algorithms, and automated systems of circulation and amplification. The model

shows how the core insights of established visual theory remain analytically productive when re-read in light of distributed agency and computational governance. In this sense, the four-dimensional framework offers a robust and adaptable architecture for analysing visual politics under contemporary conditions, without reducing political imagery to either technological determinism or representational analysis alone.

The four-dimensional framework proposed in this article is formulated as a general model of visual politics and is not limited to AI-generated or algorithmically produced imagery. In this study, empirical claims about algorithmic mediation are grounded primarily in the contemporary case of platform-circulated synthetic imagery surrounding the Pentagon misinformation episode. The historical photographs are analysed within their own media regimes to establish baseline configurations of visual politics, enabling comparison without projecting contemporary platform dynamics back onto historical cases.

Cultural Implications

For cultural analysis, the proposed model highlights how images function as sites in which meaning, memory, and collective identity are negotiated through ongoing discursive activity. Visual materials do not simply illustrate cultural values; they participate in shaping the symbolic boundaries through which communities understand belonging, trauma, responsibility, and authority. Attention to the constructive, visibility, affective, and legitimising dimensions therefore helps clarify how images operate both as aesthetic forms and as social actors within cultural discourse.

This perspective also foregrounds the interpretive labour of audiences. Viewers do not passively receive images, but actively reframe, contest, and re-embed them within existing cultural narratives and shared repertoires of meaning. Such practices of reception are central to understanding how visual significance becomes stabilised or unsettled over time, particularly as images circulate across different cultural contexts, media platforms, and communicative settings. Cultural meaning emerges through iterative processes of circulation,

interpretation, and recontextualisation that extend beyond the moment of production.

In digitally mediated cultural environments, these processes are increasingly shaped by platform infrastructures that influence how images are encountered, ranked, and rendered salient. Cultural memory, affective resonance, and symbolic visibility are increasingly organised through the combined influence of editorial practices, collective tradition, and computational systems that privilege circulation, engagement, and repetition. As a result, cultural negotiation around images unfolds within environments marked by uneven distributions of visibility and resonance, in which some visual narratives are amplified while others remain marginal.

For scholars working at the intersection of culture and communication, the framework therefore opens analytical space for examining how visuality structures intercultural understanding, cross-community dialogue, and the ethics of witnessing in digitally mediated environments. This approach situates algorithmic systems within cultural analysis, treating them as constitutive conditions under which cultural meaning, memory, and identification are produced and contested through images.

Implications for Political Analysis

From a political perspective, the proposed framework draws attention to the sensory, symbolic, and affective environments within which authority is articulated, recognised, and contested. This approach emphasises the discursive conditions through which political actors, institutions, and publics recognise certain forms of power as legitimate, intelligible, or emotionally compelling. Visual politics, in this sense, operates through the alignment of meaning, visibility, affect, and authority within shared fields of perception.

This perspective helps clarify how visual materials shape political debates and public narratives not through linear causation, but by organising the interpretive cues, emotional tonalities, and regimes of visibility through which political questions become thinkable. Images influence political outcomes by directing attention, structuring affective response, and supplying symbolic

resources that guide judgment and collective orientation. Treating visuality as a constitutive element of political life therefore enables a more nuanced understanding of how power is enacted and contested across media systems and institutional settings.

These political dynamics are increasingly conditioned by computational systems that participate in governing visibility, circulation, and response in platform-based political communication. Platform algorithms, automated curation, and recommendation systems shape which images gain prominence, the speed at which they circulate, and the affective registers they activate. Political authority and legitimacy are thus negotiated within hybrid assemblages in which human actors, institutional practices, and algorithmic processes jointly structure the conditions of recognition and response.

These developments also complicate established notions of authorship, responsibility, and accountability in political communication. When images are generated, altered, or amplified through algorithmic systems, political agency can no longer be located solely in identifiable speakers or institutions. Authority instead emerges through distributed processes of circulation and validation that may outpace deliberation, verification, or ethical judgment. The framework therefore offers analytical tools for examining how political power is exercised and contested in environments where visibility, credibility, and affect are shaped by computational governance alongside human intention.

Applicability Beyond Western Contexts

Although much foundational scholarship on visual politics has emerged from Euro-American contexts, the analytical lens developed in this article is not confined to any single cultural or political tradition. The four-dimensional framework can be adapted to political environments shaped by distinct media systems, symbolic repertoires, and historical trajectories. In many non-Western contexts, regimes of visibility are strongly influenced by state-curated imagery, ritualised performances of authority, and culturally specific norms governing emotional expression and public appearance. These settings demonstrate that the same visual mechanisms (constructing meaning, organising visibility,

mobilising affect, and articulating legitimacy) operate through different discursive logics and institutional arrangements.

As Vassileva (2021) notes, "it is dominant ideologies that shape and determine modes of discourse" (p. 73), a reminder that the interpretation of political images is always regulated by culturally embedded frameworks of meaning. Comparative analysis thus reveals the intercultural dynamics of visual politics: images that invite empathy, critique, or dissent in one context may reinforce hierarchy, unity, or national identity in another. Visual power is mediated through local histories, cultural conventions, and political expectations, producing variable effects across contexts.

A brief illustration can be found in Vietnam, where political imagery frequently foregrounds ritual, symbolism, and collective solidarity. These visual grammars of authority diverge from liberal-democratic aesthetics while remaining no less structured by visibility and affect. Read through the present framework, such practices can be analysed not as anomalies, but as culturally specific configurations of visual politics in which legitimacy is performed and recognised through distinct symbolic forms.

The era of globally circulating platforms and algorithmic infrastructures requires closer attention to how these dynamics intersect unevenly across regions. While digital systems introduce shared technical conditions of circulation, their political and cultural effects remain uneven, shaped by local media ecologies and regulatory contexts. Applying the framework across diverse settings thus highlights its flexibility as an analytical approach, enabling researchers to examine how visual politics is re-articulated at the intersection of global media technologies and culturally situated practices of meaning, visibility, and authority.

Conclusion

This article has examined the expanding significance of visual politics in contemporary public life and proposed a conceptual framework for analysing how images participate in the production of political meaning. While scholarship across visual culture, political theory, international relations, and media and

communication studies has long demonstrated that images shape what can be perceived, felt, and acted upon, these contributions have often remained analytically dispersed. By synthesising insights from Mitchell, Campbell, Hansen, Azoulay, Bleiker, and Hariman and Lucaites, the article has articulated a four--dimensional model that captures the principal mechanisms through which images become politically consequential.

The proposed framework (comprising the constructive, visibility, affective, and legitimacy dimensions) offers a systematic way to conceptualise the political agency of images. The model highlights images as active forces in political life, emphasising their capacity to structure interpretations of reality, organise regimes of visibility, mobilise collective emotions, and provide sites for the performance and recognition of political authority. The case analyses of *Saigon Execution*, *Falling Man*, and the photograph of Alan Kurdi illustrate how these dimensions operate in combination, revealing the layered nature of visual power and the diverse pathways through which images shape public sentiment, memory, and judgment.

At the same time, the article shows that these dynamics must now be understood within environments increasingly shaped by algorithmic mediation. Contemporary visual politics unfolds not only through human authorship, institutional framing, or representational content, but through hybrid assemblages in which algorithms participate in generating, circulating, amplifying, and legitimising visual forms. Under such conditions, high--velocity visibility can function as a provisional source of authority, affective resonance can be intensified through computational amplification, and questions of authorship, responsibility, and accountability become more difficult to locate.

By integrating classical visual theory with explicit attention to algorithmic mediation, the four-dimensional framework contributes to interdisciplinary debates on culture, communication, and politics while remaining adaptable across different empirical and cultural contexts. It provides a shared analytical vocabulary for examining how meaning, visibility, affect, and legitimacy are negotiated in visual cultures shaped by both human and computational agency. This perspective is particularly relevant beyond Western settings, where global

platform infrastructures intersect with locally specific media ecologies, symbolic traditions, and political imaginaries.

Future research may extend this framework to address emerging challenges such as AI-generated imagery, algorithmic regimes of visibility, digital memorialisation, and the evolving aesthetics of political leadership under conditions of synthetic media. These developments point less to the emergence of an entirely new form of visual politics than to a transformation in its operative logic. Visual politics today can be understood as a dynamic and processual field in which political meaning, authority, and responsibility are continuously negotiated at the shifting boundary between human and machine agency, as platform-mediated episodes of synthetic circulation render the conditions of visibility and legitimacy increasingly unstable.

References

Associated Press (1968). *Saigon Execution* [Photograph by Eddie Adams]. AP Images. Retrieved from https://newsroom.ap.org/editorial-photos-videos/detail?itemid=f775ebe3a9e4da11af9f0014c2589dfb&mediatype=photo. Accessed 20 May 2026.

Associated Press (2001). *Falling Man* [Photograph by Richard Drew]. AP Images. Retrieved from https://newsroom.ap.org/editorial-photos-videos/detail?itemid=347ad76a1fe5da11af9f0014c2589dfb&mediatype=photo. Accessed 20 May 2026.

Associated Press (2001). *Alan Kurdi* [Photograph by Nilufer Demir]. AP Images. Retrieved from https://newsroom.ap.org/editorial-photos-videos/detail?itemid=35138570f4454c6b9ca682a57851ecc9&mediatype=photo. Accessed 20 May 2026.

Azoulay, A. (2008). *The Civil Contract of Photography*. New York: Zone Books.

Blaney, S. (2023). Politics and Aesthetics: Jacques Rancière and Louis-Gabriel Gauny. *Philosophy & Social Criticism*, *51*(10), 1598–1618.

Bleiker, R. (2009). *Aesthetics and World Politics*. Basingstoke: Palgrave Macmillan.

Campbell, D. (2003). Cultural governance and pictorial resistance: Reflections on the imaging of war. *Review of International Studies*, *29*(S1), 57–73.

Clayton, A. (2023, May 22). *Fake AI-generated image of explosion near Pentagon spreads on social media*. The Guardian. Retrieved from https://www.theguardian.com/technology/2023/may/22/pentagon-ai-generated-image--explosion. Accessed 7 May 2026.

Hansen, L. (2015). How Images Make World Politics: International Icons and the Case of Abu Ghraib. *Review of International Studies*, *41*(2), 263–288.

Hariman, R. & Lucaites, J. L. (2016). *Public Image: Photography and Civic Spectatorship*. Chicago: University of Chicago Press.

Lehman, I., & Hoyte-West, A. (2021). Preface: Academic communication in today's world–Knowledge production and dissemination through written texts. *Discourses on Culture*, *15*(1), 1–14.

Lorés, R. (2021). From print form to digital communication: The one-way journey of academic research. *Discourses on Culture*, *15*(1), 45–63.

Mitchell, W. J. T. (1994). *Picture Theory: Essays on Verbal and Visual Representation*. Chicago: University of Chicago Press.

Vassileva, I. (2021). Rhetorical aspects of cross-cultural academic communication. *Discourses on Culture*, *15*(1), 65–81.

Zylinska, J. (2020). *AI Art: Machine Visions and Warped Dreams*. London: Open Humanities Press.