The Discursive Normalization of AI in Pedagogical Mediation and the Risks of Uncritical Acceptance of Its Interlocutions

Luís Rogério da Silva, Universidade de São Paulo, Brazil
Eliane Gonçalves, Pontifícia Universidade Católica de São Paulo, Brazil

The European Conference on Education 2025
Official Conference Proceedings

Abstract

With the increasing presence of Generative AI in professional and educational environments, this research examines the ability to distinguish between human and AI mediation in higher education, analyzing epistemological, emotional, and pedagogical impacts. Through the lens of reflexive modernity, it explores challenges to knowledge reliability, self-continuity, and emotional responsibilities displaced by simulated empathy. It also examines the erosion of epistemic trust due to plausible but imprecise AI-generated content, affecting critical knowledge construction and academic discourse. This study assesses whether the normalization of AI in pedagogical mediation fosters uncritical acceptance of its interlocutions, aligning with the "colonization of the future," where algorithmic predictions shape behaviors in opaque ways, challenging authorship, accountability, and academic integrity. Empirical research analyzed a discussion forum in a distance learning Portuguese Language course, comparing student interactions with human and AI mediators (ChatGPT-3.5). A textual corpus was evaluated by 13 human mediators based on relevance, empathy, and clarity, alongside an Adapted Reverse Turing test where AI assessed authorship. This exploratory study, using Likert-scale and open-ended questionnaires, provides insights into the ethical and pedagogical implications of AI in higher education. By addressing risks and opportunities, the findings contribute to strategies for preserving epistemic trust, fostering human-centered education, and ensuring AI enhances rather than replaces critical thinking, student engagement, and inclusive pedagogy.

Keywords: artificial intelligence, pedagogical mediation, discursive authorship, reliability, AI invisibility

iafor
The International Academic Forum
www.iafor.org

Introduction

Pedagogical mediation involves more than content transmission: it builds trust, establishes interpersonal connections, and fosters critical thinking. However, the growing presence of Generative Artificial Intelligence (GAI) in education, especially as a mediator in digital environments, introduces new challenges. One of these is the difficulty in recognizing whether mediation is conducted by humans or machines, which can undermine epistemic trust (Vallor, 2016; Zuboff, 2019) and alter the notion of academic authorship.

This study discusses the concept of simulated empathy, understood as the GAI's ability to deliver emotionally charged responses without subjective experience. Such simulation can lead students to perceive genuine empathy where there is only programming, making it harder to distinguish between human and automated support. According to Elish (2019), these interactions place users in a "moral crumple zone," shifting responsibility for GAI failures to human operators. Crawford (2021) notes that this false empathy can create affective bonds with AI, compromising students' intellectual autonomy and weakening critical capacity (Danaher, 2019; Duah & McGivern, 2023).

The study investigates the normalization of GAI in pedagogical mediation, understood as the process by which its use becomes routine and unquestioned. This naturalization promotes the invisibility of AI, obscuring authorship and making it difficult to distinguish between human and machine. This invisibility relates to Ulrich Beck's (2010) concept of "colonization of the future," in which educational decisions become shaped by invisible and decontextualized algorithmic predictions (Eubanks, 2018; Pasquale, 2020).

The research is framed within the theoretical context of Risk Society, as outlined by Beck (2010) and Giddens (1991). Both highlight how trust in abstract technological systems, increasingly ubiquitous, can mask structural risks. The reliance on invisible AI mediators in virtual learning environments is a clear example of reflexive modernity. As Giddens (1991) states, specialized systems are used without full understanding by individuals, transferring trust from users to specialists and programmers.

Zuboff (2016, 2019) reinforces this analysis with the concept of Surveillance Capitalism, in which behavioral data extracted from digital interactions are turned into predictive products. In education, AI mediators in distance learning forums silently collect data, influencing the behavior of both students and teachers. Thus, AI mediation is not neutral: it structures interactions in ways that feed prediction models and algorithmic learning, creating an opaque surveillance environment.

These dynamics are also expressed discursively. The simulated empathy of GAI is not merely functional—it shapes the content and pedagogical meaning of the interaction. These machines employ strategies of active listening and emotional support that reinforce the illusion of affective interlocution (Crawford, 2021). Discursively, these strategies replicate internalized social norms, increasing uncritical acceptance of AI as a legitimate agent in educational spaces.

Elish (2019) warns of the ethical risk of responsibility transfer: when AI fails in mediation, the blame may fall on the human teacher, who does not control the algorithms. The construction of an artificial ethos—i.e., a discursive image of competence and empathy—sustains this illusion. As Danaher (2019) and Zuboff (2019) show, this trust is often

unjustified, since generative models can produce imprecise or misleading content with high plausibility.

To mitigate the negative impacts of simulated empathy, authors such as Pasquale (2020) and Eubanks (2018) advocate for transparency policies and AI regulation in education. Clear signaling of AI presence, restrictions on its use in contexts requiring real emotional support, and digital literacy for students are essential strategies.

GAI-generated discourse can also be examined through enunciation theory and discourse analysis as proposed by Maingueneau (2008, 2013, 2015). The author introduces the concepts of scenography and ethos, showing how GAI language simulates a legitimate speaker, shaping perceptions of trustworthiness. This discursive construction is especially relevant in educational contexts, where mediation presupposes authenticity and active listening. AI projects an ethos of support and competence, even without subjectivity, which favors its acceptance as a mediator.

The increasing sophistication of AI-generated language makes authorship identification more difficult. Researchers at the University of Maryland (2023) found that AI text detectors can be easily deceived by paraphrasing. OpenAI itself states that its tool is only 26% accurate, with a 9% false-positive rate (McCrosky, 2024). Reeves and Nass (1996) also demonstrate that these algorithms are biased against non-native speakers, compounding issues of epistemic justice.

Authorship is a central issue. Gunkel (2023) asserts that writing in AI-mediated environments is characterized by distributed authorship, shared between humans and machines. Fernandes (2023) refers to a "simulacrum of author function," while Gallo (2023) questions whether AI textual production constitutes genuine authorship or mere statistical reproduction. Simulated empathy, by intensifying the illusion of authenticity, deepens this dilemma and reinforces the attribution of subjectivity to systems that, in fact, lack it.

In the field of genre analysis, Bakhtin (1997) and Marcuschi (2008) emphasize that educational forum interactions follow relatively stable structures. Even in this context, AI adapts its discourse to prevailing genres, establishing itself as a legitimate part of the environment. By adhering to these conventions, GAI not only imitates but participates in the construction of pedagogical and collaborative meaning.

Statement of the Problem

The growing presence of Generative Artificial Intelligence (GAI) in educational settings, especially in pedagogical mediation within distance learning courses, raises complex challenges related to the identification of authorship and its epistemological and emotional implications. The difficulty in distinguishing whether mediation is conducted by human agents or artificially intelligent agents threatens epistemic trust, a fundamental element for knowledge construction and the development of students' intellectual autonomy.

Furthermore, GAI's simulation of empathy—an emotionally charged but programmed response without true subjectivity—can create a false impression of genuine emotional support, compromising students' critical capacity. This phenomenon, coupled with the invisibility of artificial intelligence in educational environments, hinders clear authorship identification and contributes to the naturalization of AI use in pedagogical contexts. This

process can be understood as a “colonization of the future,” in which decisions and behaviors are shaped by opaque algorithmic predictions.

In this scenario, there is also an ethical risk related to responsibility transfer, where failures in AI-mediated mediation are unfairly attributed to human instructors or operators, reinforcing an artificial ethos that legitimizes AI as a competent and empathetic interlocutor. Additionally, the production of content that is plausible but imprecise or erroneous challenges the construction of reliable knowledge, demanding reflection on transparency, regulation, and digital literacy.

Therefore, the central problem of this research is to investigate how the invisible presence of GAI in pedagogical mediation impacts perceptions of discursive authorship, epistemic trust, and emotional dynamics in higher education environments—particularly analyzing the risks of uncritical acceptance and the consequences for ethics and educational quality mediated by intelligent technologies.

Study Objectives

This study aims to investigate the growing presence and normalization of Generative Artificial Intelligence (GAI) as a mediator in higher education, focusing on its epistemological and emotional impacts. The specific objectives are:

1. To examine the ability of students and educators to distinguish between mediation conducted by humans and by artificially intelligent agents in digital learning environments.
2. To assess how the invisibility and normalization of AI in pedagogical mediation influence perceptions of authorship and trustworthiness.
3. To empirically analyze the discourse of a distance learning forum, contrasting human and AI-mediated interactions through qualitative and quantitative methods, including evaluations by human evaluators and AI authorship detection tests.

Research Questions

How effectively can students and educators identify whether pedagogical mediation is conducted by humans or by Generative Artificial Intelligence in digital higher education environments?

1. What are the epistemological and emotional consequences of AI-simulated empathy on students' learning experiences and critical thinking skills?
2. How do the invisibility and normalization of AI mediators affect epistemic trust and perceptions of authorship?
3. How do human evaluators and AI authorship detection tools perform in distinguishing between AI-generated and human-generated mediation texts?

Research Hypotheses

H1: Students and educators have significant difficulty distinguishing between human-led and generative AI-driven mediation in digital learning forums.

H2: Empathy simulation by AI mediators creates affective engagement that can compromise students' intellectual autonomy and critical thinking skills.

H3: The normalization and invisibility of AI in pedagogical mediation leads to reduced epistemic trust and obscured authorship attribution.

H4: Responsibility for failures in AI-mediated pedagogical interactions is implicitly transferred to human educators, despite their lack of control over the AI.
H5: Generative AI adapts its discourse according to prevailing educational genres, thus reinforcing its acceptance and legitimacy as a pedagogical mediator.

Methodology

This exploratory research focuses on the difficulty of distinguishing human from artificial authorship in the context of pedagogical mediation in distance education environments. Based on the use of Generative Artificial Intelligence (GAI), especially the ChatGPT 3.5 model, the study questions the growing presence of AI as a mediator in educational forums, using a theoretical framework including Turing (1950), Gunkel (2023), Beck (2010), Giddens (1991), Zuboff (2019), Elish (2019), and others. The research investigates the limits of human and algorithmic perception regarding authorship of AI-mediated discourse, analyzing its epistemic, emotional, and ethical impacts.

The methodology, structured in eight stages, involved collecting real excerpts from forums in language courses and creating two distinct corpora: one consisting of interactions between students and human mediators and the other between students and GAI. A third mixed corpus was created from these. The data were evaluated by thirteen human mediators who assessed the quality and assumed authorship of the interactions. Simultaneously, GAI was tasked with performing the same evaluation using what was called the Adapted Reverse Turing Test. A third verification tool, GPTZero, was also employed to detect authorship based on linguistic and stylistic patterns.

In each evaluated interaction, humans were asked to assess mediation in terms of relevance, empathy, and clarity. The AI was given the same task before being asked to infer the nature of the responding agent. The results showed that in most analyzed interactions, both humans and AI classified GAI responses as being authored by humans. For example, in Interaction 1, about 70% of experts rated the mediation as “good” or “excellent,” and 80% believed it was human. GAI, in turn, rated the quality as “good,” likely of human authorship. Similar results were found in subsequent interactions, including those with more affective mediation or controversial content, such as Interaction 4, where AI responded empathetically to a harsh critique from a student.

Analysis of Results

These results demonstrate a concerning pattern of misattributed authorship, supported by two central phenomena: the “normalization” of GAI and its “invisibility” in educational environments. Statistical analysis of the results revealed that in both the traditional and Adapted Reverse Turing Tests, the correct detection rate was only 20%. That is, in 80% of cases, GAI was mistaken for a human. This finding supports Floridi and Chiriatti’s (2020) argument about the obsolescence of the Turing Test as a reliable measure of artificial intelligence in contemporary contexts. Moreover, it shows that even trained human interlocutors face significant limitations in detecting the nature of responses in digitally mediated environments.

The results also highlight that GAIs, by simulating empathy with discursive fluency, can deceive both human evaluators and themselves. This supports concerns raised by authors like Madeleine Clare Elish (2019), who warns about the “moral crumple zone,” in which

algorithmic simulation shifts emotional and moral responsibility to human operators. AI thus appears not merely as a neutral tool but as a discursive agent constructing its own ethos—a term that, according to Maingueneau (2008), refers to the image of trustworthiness projected in discourse. AI formulates responses with politeness, active listening, and motivational suggestions that mimic the actions of a caring and attentive educator, despite lacking subjectivity.

This simulation capacity becomes even more evident when AI is evaluated in terms of empathy. As Crawford (2021) and Danaher (2019) demonstrate, empathy simulated by algorithms reinforces uncritical acceptance of their responses, especially in educational settings where emotional support is expected. The difficulty in distinguishing between human and artificial empathy weakens epistemic trust, as programmed responses are perceived as authentic interactions. This reflects Ulrich Beck's (2010) notion of "colonization of the future," which warns of technologies shaping human behaviors through algorithmic predictions without transparency or critical debate.

In this scenario, the sophistication of language models emerges as a significant challenge for contemporary pedagogy. The fact that GAIs can respond with coherence, fluency, and artificial empathy means that the "humanity" of discourse can no longer be assessed intuitively. Recent studies from the University of Maryland and OpenAI (2023) point out high error rates in detecting artificial authorship, with false positive and negative rates compromising trust in automated detectors. GPTZero, for instance, failed to correctly analyze part of the samples tested in this research.

Furthermore, the analysis shows that educational forums constitute their own discursive genres, with rules and conventions internalized by participants, as noted by Bakhtin (1997) and Marcuschi (2008). Within this context, AI learns to model its utterances according to these rules, reinforcing its discursive camouflage. David Gunkel's (2023) notion of "distributed authorship" gains strength, as it demonstrates that discourse produced in digital environments is no longer the result of a single subjectivity, but a blending of human and machine.

Conclusion

The results indicate that the trust attributed to AI is anchored in its discursive performance. Giddens (1991) observes that in late modernity, individuals place trust in abstract systems they do not understand, granting them authority based on the credibility of their developers. GAI, by adopting a polite tone and presenting organized information, builds its ethos as a legitimate mediator, even without clearly disclosing that it is not human. This uncritical acceptance is amplified by daily familiarity with AI tools and the lack of explicit authorship markers in educational interfaces, as discussed by Selwyn (2019).

The research thus demonstrates that the presence of GAI as an invisible and indistinguishable pedagogical mediator poses ethical, pedagogical, and epistemological challenges. The successful simulation of empathetic mediation masks the absence of subjectivity and shifts responsibility, demanding better preparation from human mediators and the development of protocols that clearly indicate the authorship of interactions. As Pasquale (2020) and Eubanks (2018) point out, it is necessary to establish regulations to ensure that AI is used transparently and not in contexts that require genuine human mediation.

Given the evidence that both humans and GAIs achieve only 20% accuracy in identifying authorship, it is urgent to promote algorithmic literacy among students and teachers, enabling them to recognize signs of AI presence and critically question the discourses it produces. The absence of such literacy can compromise the quality of pedagogical mediation and learners' autonomy, while reinforcing AI's “invisibility” in the educational ecosystem.

Finally, the research confirms that pedagogical mediation cannot be reduced to the production of plausible responses; it requires genuine empathy, subjective responsibility, and the collective construction of knowledge. However advanced GAI may be, it still operates in the realm of simulation, without lived experience or emotional commitment. Therefore, its presence in education must be carefully regulated to preserve authentic human interaction as the foundation of epistemic trust and critical formation.

References

Bakhtin, M. (1997). *Aesthetics of Verbal Creation* (2nd ed.). São Paulo: Martins Fontes.

Beck, U. (2010). *Risk Society: Towards a New Modernity*. São Paulo: Editora 34.

Bender, E. M., Gebru, T., McMillan-Major, A., & Mitchell, M. (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. https://doi.org/10.1145/3442188.3445922

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. New Haven: Yale University Press.

Danaher, J. (2019). *Automation and Utopia: Human flourishing in a world without work.* Cambridge: Harvard University Press.

Duah, J. E., & Mc Givern, P. (2024). *How generative artificial intelligence has blurred notions of authorial identity and academic norms in higher education, necessitating clear university usage policies*. International Journal of Information and Learning Technology, 41(2), 180–193. https://doi.org/10.1108/IJILT-11-2023-0213

Elish, M. C. (2019). *Moral crumple zones: Cautionary tales in human-robot interaction*. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. https://doi.org/10.1145/3306618.3314281

Eubanks, V. (2018). *Automating Inequality: How high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press.

Fernandes, C. (2023). *Authorship in texts produced by artificial intelligence and by students from a discursive perspective*. Revista da ABRALIN, 22(1), 45–60. https://revista.abralin.org/index.php/abralin/article/view/2183/2805

Floridi, L., & Chiriatti, M. (2020). *GPT-3: Its nature, scope, limits, and consequences*. Minds and Machines, 30, 681–694. https://doi.org/10.1007/s11023-020-09548-1

Gallo, S. M. L. (2023). *ChatGPT: hyperauthor or not author?* Linguistic Traits – Journal of Linguistic Studies, 7(1), 84–95. https://periodicos.unemat.br/index.php/tracos/article/view/11199

Gehrmann, S., Strobel, B., Bastianello, M., & Stickland, A. (2019). *GLTR: Statistical detection and visualization of generated text*. arXiv preprint arXiv:1906.04043. https://arxiv.org/abs/1906.04043

Giddens, A. (1991). The Consequences of Modernity. São Paulo: Editora Unesp.

Gunkel, D. J. (2023). *Person, Thing, Robot: A Moral and Legal Ontology for the 21st Century and Beyond*. MIT Press.

Maingueneau, D. (2008). *Enunciation Scenes*. São Paulo: Parábola Editorial.

Maingueneau, D. (2013). *Genesis of Discourses*. São Paulo: Parábola Editorial.

Maingueneau, D. (2015). *Discourse and Discourse Analysis*. São Paulo: Parábola Editorial.

Maingueneau, D. (2021). *The Margins of Discourse*. São Paulo: Editora Contexto.

Marcuschi, L. A. (2008). *Textual Genres: Definition and Functionality*. In Â. P. Dionísio, A. R. Machado, & M. A. Bezerra (Eds.), Textual Genres & Teaching (6th ed., pp. 19–36). Rio de Janeiro: Lucerna.

McCrosky, J. (2024, March 28). *Who wrote that? Evaluating tools to detect AI-generated text*. Mozilla Foundation. https://foundation.mozilla.org/pt-BR/blog/who-wrote-that-evaluating-tools-to-detect-ai-generated-text/

Miller, T. (2019). *Explanation in artificial intelligence: Insights from the social sciences*. Artificial Intelligence, 267, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Pasquale, F. (2020). *The Black Box Society: The secret algorithms that control money and information*. Cambridge: Harvard University Press.

Reeves, B., & Nass, C. (1996). *The Media Equation: How people treat computers, television, and new media like real people and places*. Cambridge: Cambridge University Press.

Röhe, A., & Santaella, L. (2023). *Confusions and dilemmas of anthropomorphizing artificial intelligences*. Revista PUC-SP, 30(2), 112–125. https://revistas.pucsp.br/index.php/teccogs/article/view/67070/45078

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Social Science Research Network (SSRN). https://doi.org/10.2139/ssrn.3338750

Turing, A. M. (1950). *Computing machinery and intelligence*. Mind, 59(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

University of Maryland. (2023). *AI-generated content: Is it actually detectable?* College of Computer, Mathematical, and Natural Sciences. https://cmns.umd.edu/news-events/news/ai-generated-content-actually-detectable

Vallor, S. (2016). Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting. Oxford: Oxford University Press.

Zuboff, S. (2016). *Secrets of surveillance capital*. Frankfurter Allgemeine Zeitung. https://www.faz.net/aktuell/feuilleton/debatten/the-digital-debate/shoshana-zuboff-secrets-of-surveillance-capital-14103616.html

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The fight for a human future at the new frontier of power*. New York: PublicAffairs.