Feeling Informed but Remembering Less: The Effect of AI Attribution on Learning Effectiveness, Trust, and Acceptance of Educational Videos

Marcus Athmer[1] and Caterina Fox[2]
[1]IU Internationale Hochschule, Merzen, Germany
[2]IU Internationale Hochschule, Mannheim, Germany

marcus.athmer@gmail.com
caterina.fox@iu.org

Abstract: Generative artificial intelligence (AI) is increasingly used to create educational materials, including instructional videos. This study investigates whether labeling educational content as AI-generated versus human-created influences learning outcomes. Using a randomized online A/B experiment, 211 participants were exposed to the same educational video under either an AI-generated or human-created label. The findings reveal a differentiated pattern: AI attribution did not significantly affect learners' subjective perceptions of understanding or informativeness but was associated with lower factual recall. Furthermore, AI-labeled videos resulted in lower trust, higher perceived risk, and reduced intentions to recommend or share the content. The results suggest that AI labels function as heuristic cues that influence how learners evaluate educational materials and may introduce additional cognitive demands during learning. The study contributes to research on AI-supported education by demonstrating that the perceived origin of instructional content can shape learning experiences beyond the characteristics of the content itself.

Keywords: Artificial intelligence (AI), AI Attribution, Educational Videos, Learning Effectiveness, Trust in AI

1. Introduction

Advances in generative AI have enabled efficient creation of high-quality instructional videos that can be used in higher education, corporate training, and online learning. These materials offer opportunities for personalized learning, scalable content production, multilingual delivery, and rapid material creation (Grewal et al, 2024). However, less is known about how learners respond when AI-generated content is explicitly labeled as such (Lin and Zhang 2026). Emerging evidence suggests that AI labels can function as heuristic cues that shape perceptions of credibility, authenticity, and trust independently of the actual quality of the content (Koning and Voorveld, 2025; Schilke and Reimann, 2025). Consequently, learners may evaluate identical educational materials differently solely because of their attributed origin. However, previous research has rarely distinguished between the objective quality of AI-generated content and the psychological consequences of AI attribution alone (Lim and Schmälzle, 2024). This study separates attribution from content quality and thus contributes to a more nuanced understanding of how AI disclosure affects the reception of educational videos and offers implications for the design and communication of AI-assisted learning materials.

2. Theoretical Foundation

2.1 AI Attribution as a Heuristic Cue and Learning Effectiveness of Educational Videos

It has become increasingly difficult for learners to distinguish between human- and AI-generated materials based on the content alone. Instead, they frequently rely on contextual information such as origin labels, which function as heuristic cues that shape information processing (Nguyen et al, 2026; Sundar and Kim, 2019; Yang and Sundar, 2024). An AI label may activate existing beliefs about artificial intelligence and thus reduce perceived authenticity and credibility of the material. In this way, the label itself becomes part of the learning experience (Mariadassou et al, 2024). From the perspective of Cognitive Load Theory (Mayer, 2005), such attribution effects may influence the allocation of cognitive resources during learning. If learners associate AI-generated content with uncertainty, reduced reliability, or lower authenticity, they may devote additional cognitive effort to evaluating the source rather than processing the instructional information itself. This additional cognitive load may interfere with knowledge acquisition and retention, even when the underlying learning materials remain unchanged. This leads to the first hypothesis:

H1: AI attribution reduces learning effectiveness compared to identical videos labeled as human-created.

2.2 Trust, Perceived Risk, and Acceptance

To be effective, learners must trust the educational materials they are using. However, because AI is often associated with misinformation, hallucinations, and a lack of human oversight, AI attribution can increase perceived risk (Baek et al, 2026). Thus, perceived risk represents an important psychological mechanism linking AI attribution to trust and subsequent behavioral responses (Liu and Lu, 2026). Lower trust is likely to reduce learners' willingness to engage with AI-labeled educational materials, recommend them to others, or rely on them for future learning. Building on these theoretical considerations, hypotheses 2-4 were developed:

H2: AI attribution reduces trust compared to identical videos labeled as human-created.

H3: AI attribution increases perceived risk compared to identical videos labeled as human-created.

H4: AI attribution reduces sharing intentions compared to identical videos labeled as human-created.

3. Methodology

To test the hypotheses, a randomized A/B online survey experiment was conducted. Participants were randomly assigned to one of two experimental groups. In the AI-label condition, participants were informed that the educational video had been generated using artificial intelligence, whereas participants in the human-label condition were informed that the same video, a 51-second baking tutorial (iStock video), had been created by humans. Thus, the only manipulated variable was the perceived origin of the content. Following video exposure, participants completed measures assessing learning effectiveness, trust, perceived risk, and behavioral intentions using seven-point Likert scales, ranging from 1 (“strongly disagree”) to 7 (“strongly agree”). In addition to subjective evaluations, an open-ended recall task was included to assess objective learning outcomes. Participants were asked to recall factual information presented in the video without being provided with response options. The number of correctly recalled details served as an indicator of actual retention. Manipulation checks were conducted to verify whether participants recognized the assigned attribution condition.

The final sample consisted of 211 participants (107 in the AI-label condition, 104 in the human-label condition), recruited through a convenience self-selection procedure, with random assignment ensuring comparable group sizes and reducing the likelihood of systematic differences between conditions. Welch’s t-tests were employed to examine differences between the AI-label and human-label conditions across the outcome variables (Delacre et al, 2022).

4. Findings

The manipulation check confirmed that participants perceived the intended difference between the two experimental conditions. Although both groups viewed the identical educational video, participants in the AI-labeled condition were significantly more likely to perceive the video as AI-generated, indicating that the attribution manipulation was successful.

4.1 Effects of AI Attribution on Learning Effectiveness

The results revealed a differentiated pattern for learning effectiveness. Contrary to H1, AI attribution did not significantly affect participants' self-reported learning. Respondents in both conditions rated the educational video as equally understandable and informative. However, objective recall told a different story. Participants who believed the video had been created by a human recalled more factual information than those who viewed the identical video labeled as AI-generated (Welch-$t(207.42) = 2.21$; $p = .028$; $d = 0.30$). This suggests that AI attribution influenced memory performance despite stable subjective evaluations of learning.

4.2 The Role of Perceived Risk and Trust

Participants exposed to the AI label reported significantly lower trust (Welch-$t(208.02) = 4.01$; $p < .001$; $d = 0.55$) in the educational video, supporting H2. AI attribution also increased video-specific perceived risk (Welch-$t(208.99) = -2.42$; $p = .016$; $d = -0.33$), and lowered intentions to recommend (Welch-$t(207.03) = 2.93$; $p = .004$; $d = 0.40$) or share the video (Welch-$t(208.91) = 3.36$; $p < .001$; $d = 0.46$), supporting H3 and H4 respectively.

Importantly, higher perceived risks were consistently associated with lower trust and weaker behavioral intentions. Participants who perceived greater risks evaluated the educational video less positively and were

less willing to engage with or recommend it. Trust emerged as a key determinant of learners' acceptance. Higher levels of trust were associated with stronger intentions to use, recommend, and engage with educational videos. Together, these findings indicate that the effects of AI attribution operate primarily through affective evaluations and perceived risk rather than through changes in learners' subjective perceptions of understanding.

5. Discussion

Overall, the findings suggest that labeling educational content as AI-generated does not substantially alter how informative learners believe the content to be, but it reduces trust, lowers behavioral intentions, and is associated with poorer recall of instructional details. This pattern supports the notion that origin labels act as heuristic cues, influencing judgments through pre-existing beliefs about artificial intelligence.

A particularly noteworthy finding is the discrepancy between perceived and objective learning. Participants who viewed the AI-labeled video reported learning experiences comparable to those in the human-labeled condition but recalled significantly less factual information. This divergence suggests that subjective learning evaluations are not necessarily reliable indicators of actual knowledge acquisition when AI attribution is involved.

The findings also underline the importance of trust as a central mechanism linking AI attribution to learner acceptance. Rather than rejecting AI-generated educational content outright, participants appeared to evaluate it more cautiously, suggesting that concerns about reliability, authenticity, and accountability continue to shape attitudes toward AI-supported learning.

From a practical perspective, the results suggest that educational institutions and content providers should carefully consider how AI disclosure is communicated. While transparency regarding AI use is essential from an ethical standpoint, disclosure alone may unintentionally reduce learner trust and engagement if existing concerns about AI remain unaddressed. Complementing AI attribution with information about human oversight, quality assurance, or expert review may help mitigate these effects.

This study is subject to several limitations, including the use of a single educational video and the measurement of immediate learning outcomes within one experimental session.

Overall, this study demonstrates that learners do not evaluate educational videos solely on the basis of their content. Simply believing that instructional material was created by AI can alter trust, acceptance, and even memory performance, underscoring the importance of considering attribution effects as AI-generated educational resources become increasingly prevalent.

6. Conclusion

As generative AI becomes increasingly integrated into educational content production, understanding how learners respond to AI-generated materials is becoming as important as evaluating the quality of the materials themselves. The findings of this study suggest that AI labels function as contextual signals that shift learners' attention toward evaluating trustworthiness and potential risks. This additional evaluative processing may consume cognitive resources that would otherwise be allocated to content processing, resulting in lower objective recall despite unchanged subjective perceptions of learning. For educational practice, these results imply that learners may feel they have understood and retained information equally well when consuming AI-generated instructional content, while their actual retention of details may be diminished.

The study contributes to the growing literature on AI in education by distinguishing the psychological effects of AI attribution from the quality of AI-generated content. As institutions increasingly adopt transparent disclosure practices for AI-assisted learning materials, understanding how such disclosures influence learner perceptions and outcomes becomes critical. Future research should examine whether attribution effects persist across different subject areas, learner populations, and longer-term retention intervals. Moreover, investigating the psychological mechanisms underlying reduced recall—such as changes in attention, motivation, or cognitive engagement—would further advance understanding of how AI attribution shapes learning processes. Ensuring that AI-supported educational content is not only accurate but also perceived as credible and conducive to learning will be essential for realizing the full potential of generative AI in education.

Ethics Declaration

Ethical clearance was not required for the research. The study adhered to ethical guidelines for research involving human participants. Participants were informed about the study's general purpose, participants' rights, and ethical guidelines.

AI Declaration

ChatGPT was used to improve flow and academic tone in several sections of this paper.

References

Baek, T.H., Kim, J. and Kim, J.H. (2026) "Effect of Disclosing AI-Generated Content on Prosocial Advertising Evaluation", *International Journal of Advertising*, Vol 45, No. 1, pp 171–192. https://doi.org/10.1080/02650487.2024.2401319

Delacre, M., Lakens, D. and Leys, C. (2022) "Correction: Why Psychologists Should by Default Use Welch's t-test Instead of Student's t-test", *International Review of Social Psychology*, Vol 35, No. 1, pp 21–21. https://doi.org/10.5334/irsp.661

Grewal, D., Satornino, C.B., Davenport, T. and Guha, A. (2025) "How Generative AI is Shaping the Future of Marketing", *Journal of the Academy of Marketing Science*, Vol 53, No. 3, pp 702–722. https://doi.org/10.1007/s11747-024-01064-3

Koning, B. and Voorveld, H.A. (2025) "Disclaimer! This Content is AI-Generated: How AI-Disclosures Influence Trust in Advertisements and Organizations", *Journal of Interactive Advertising*, Vol 25, No. 3, pp 240–253. https://doi.org/10.1080/15252019.2025.2554149

Lim, S. and Schmälzle, R. (2024) "The Effect of Source Disclosure on Evaluation of AI-Generated Messages", *Computers in Human Behavior: Artificial Humans*, Vol 2, No. 1, 100058. https://doi.org/10.1016/j.chbah.2024.100058

Lin, T. and Zhang, Y. (2026) "Visible Sources and Invisible Risks: Exploring the Impact of AI Disclosure on Perceived Credibility of AI-Generated Content", *Journal of Science Communication*, Vol 25, No. 1. https://doi.org/10.22323/358020260107085703

Liu, J. and Lu, L. (2026) "How AI-Generated Content Shapes User Trust: The Roles of Cognitive Processing, Perceived Risk, and Transparency Labels", *Behavioral Sciences*, Vol 16, No. 6, 957. https://doi.org/10.3390/bs16060957

Mariadassou, S., Klesse, A.K. and Boegershausen, J. (2024) "Averse to What: Consumer Aversion to Algorithmic Labels, but Not Their Outputs?", *Current Opinion in Psychology*, Vol 58, 101839. https://doi.org/10.1016/j.copsyc.2024.101839

Mayer, R.E. (2005) *The Cambridge Handbook of Multimedia Learning*, Cambridge University Press.

Nguyen, H.P., Le, P.U. and Le, X.T. (2026) "From AI Disclosure to Purchase Intention: The Mediating Role of Trust and the Moderating Effect of Collectivistic Orientation in Vietnam", *Innovative Marketing*, Vol 22, No. 1, pp 274–284. http://dx.doi.org/10.21511/im.22(1).2026.20

Schilke, O. and Reimann, M. (2025) "The Transparency Dilemma: How AI Disclosure Erodes Trust", *Organizational Behavior and Human Decision Processes*, Vol 188, 104405. https://doi.org/10.1016/j.obhdp.2025.104405

Sundar, S.S. and Kim, J. (2019) "Machine Heuristic: When We Trust Computers More Than Humans with Our Personal Information", Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, May 4–9, Glasgow, Scotland, UK, paper 538, pp 1–9. https://doi.org/10.1145/3290605.3300768

Yang, H. and Sundar, S.S. (2024) "Machine Heuristic: Concept Explication and Development of a Measurement Scale", *Journal of Computer-Mediated Communication*, Vol 29, No. 6, zmae019. https://doi.org/10.1093/jcmc/zmae019