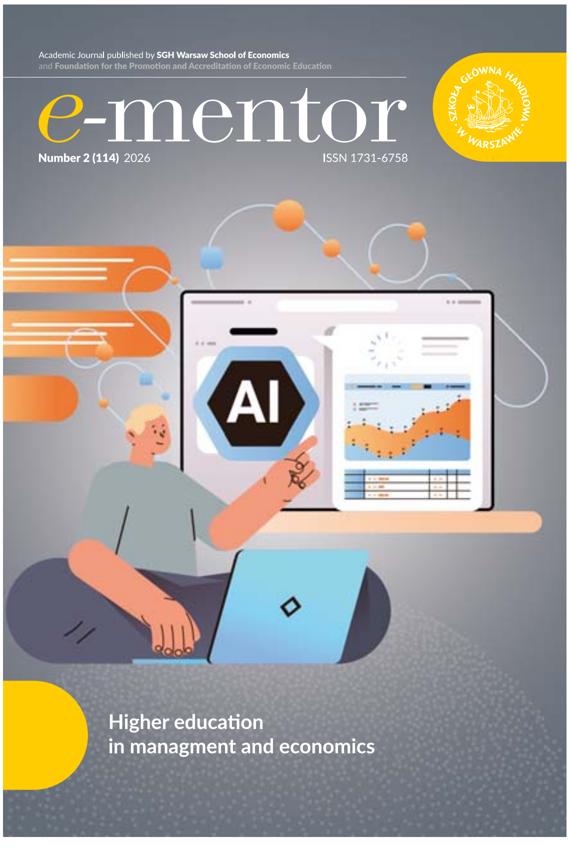
e-mentor

DWUMIESIĘCZNIK SZKOŁY GŁÓWNEJ HANDLOWEJ W WARSZAWIE
WSPÓŁWYDAWCA: FUNDACJA PROMOCJI I AKREDYTACJ KIERUNKÓW EKONOMICZNYCH



2026, nr 2 (114)

Božić Lenard, D., & Kic-Drgas, J. (2026). Clarity and credibility: a comparative study of students' perception of AI-Generated and human-created content. *e-mentor, 2*(114), 14–22. https://www.doi.org/10.15219/em114.1746

Dragana Božić Lenard

Joanna Kic-Drgas

Clarity and Credibility: A Comparative Study of Students' Perception of AI-Generated and Human-Created Content

Abstract

The growing popularity of AI across various fields affects students' decisions. Since one of the modalities of AI is clarification and explanation of more complex phenomena, the focus of this paper is on how knowledge processing differs between AI- and human-generated texts, which texts readers trust more, and why. This study investigates how university students perceive the clarity and credibility of explanations produced by AI in comparison to those written by humans. Using a quasi-experimental, mixed-methods design, 298 students from universities in Croatia and Poland participated in a survey assessing their ability to recognise the source of short academic texts and evaluate their reliability, usefulness, and comprehensibility. The results reveal a high level of AI adoption among students, as well as an emerging ability to distinguish between AI- and human-authored content based on linguistic and stylistic cues. Notably, students who reported frequent AI use were more confident in their evaluations and more trusting of AI-generated explanations. Disciplinary background, gender and topic familiarity also influenced content preferences and recognition accuracy. Across both logic and neurobiology tasks, students showed a marked preference for hybrid explanations that combine AI-generated precision with human-authored contextual clarity. These findings contribute to the growing field of AI literacy and offer practical implications for the design of trustworthy, pedagogically effective AI-supported learning materials.

Keywords: clarity, credibility, AI-generated content, human-created content, student perceptions

Introduction

Considering the changes (technological, social, demographical etc.) ongoing in the current world, the understanding of diverse concepts from various disciplines seems a crucial element of living and actively participating in society. Explanatory elements are observable not only in everyday life but also in professional fields. This is why understanding is also key for decision-making processes and for shaping critical reasoning and perceptions of the surrounding world.

Just as they are applied in other disciplines, explanations also play an important role in the learning and teaching process. Constructing a valid explanation is not solely a framework for understanding; it also includes many factors that contribute to its clarity and validity, such as logical structure, exemplifications and transparent cause-and-effect thinking (Moore & Wright, 2023).

One of the most transformative developments in this regard is the emergence of generative artificial intelligence (AI). AI-powered tools are increasingly capable of producing coherent, structured, and seemingly authoritative explanations across a range of topics. As these tools become embedded in educational environments, they raise important questions about how students perceive and evaluate the clarity, reliability, and trustworthiness of AI-generated versus human-created content. Human-computer interaction in this context is not only a matter of usability or interface design but also a cognitive and epistemological concern: how do learners distinguish machine-gener-

Dragana Božić Lenard, University of Osijek, Croatia, https://orcid.org/0000-0003-4479-1329
Joanna Kic-Drgas, Adam Mickiewicz University, Poland, https://orcid.org/0000-0002-8133-9190

ated from human-authored explanations, and how does this influence their trust and learning?

This study aims to examine students' ability to critically evaluate the credibility of AI-generated content, a skill that may prove crucial in the future. The primary objective of this research is to determine whether students possess sufficient linguistic and critical awareness to distinguish between AI-generated and human-authored texts, thereby enabling informed and responsible content consumption.

Literature Review

Explanations – Definition, Types and Structure

Every explanation is deeply connected to the way people perceive and interpret the world. The concept of explanation itself is multidimensional, and its definition has long posed challenges for researchers. An explanation can be defined in various ways depending on the perspective adopted, leading to different interpretations and misconceptions (Tan, 2022). Josephson and Josephson (1996) define explanations as 'assessments of causal responsibility,' thereby situating explanation within the domain of cause-and-effect reasoning.

Considering the latter definition, Miller (2019) states that an explanation is based on bridging the gap between the explanandum and explanans. An explanation can also be understood as the result of a cognitive process and, as such, can be perceived as a product. Finally, it is also a result of 'knowledge transfer' between an explainer and explainee (Miller, 2019). Poggiolessi (2024) identifies causal explanations as those that illuminate the causes of phenomena, in opposition to conceptual explanations, which focus on the reasons for those phenomena. Following the terminology introduced by Leinhardt and Steele (2005), explanations used in educational contexts are instructional explanations; they serve the cognitive needs of learners (Keil & Wilson, 2000). Moore and Wright (2023) shift the perspective and strongly emphasise the importance of a well-posed question in the explanatory process, rather than the type of explanation itself; however, explanations should answer the questions of how and why, rather than what (Federer et al., 2015).

Human Explanations – Features

Miller (2019) discusses the characteristics of human explanations based on an analysis of recent literature on the topic. What distinguishes human explanations is their contrastive nature, which highlights the phenomenon being explained through contrast (Woodward, 2005). Furthermore, human explanations are inherently biased, as they rely on selecting the most likely option. Although reality and probability are crucial for explaining phenomena, for humans, they are less valuable than presenting the causes of a given situation or phenomenon. Finally, explanations are based on knowledge construction, which means they are dependent on the explainer's and explainee's belief systems (Miller, 2019). Lombrozo and Liquin (2023) highlight the selectivity and incompleteness of human explanations as some of their greatest strengths. Selectivity responds to epistemic curiosity, enabling individuals to process new information effectively and integrate it into existing cognitive structures. Moreover, selectivity often reflects an individual's motivations, purposes, and intentions (Kim et al., 2023).

However, a lack of clarity and precision in explanations may reduce their perceived credibility and trustworthiness. This issue is particularly relevant in the context of artificial intelligence, where users often rely on AI-generated explanations to understand complex information. When such explanations are vague, ambiguous, or insufficiently transparent, they can undermine users' trust in AI systems and reduce their willingness to accept or act on the information provided (Chan & Hun, 2023; Pitts et al., 2025). An experiment conducted by Rago et al. (2024) confirmed a correlation between the format and clarity of an explanation and the level of trust in it; full experimental details are provided in Appendix A.

AI-Generated Explanations

The scope of AI implementation continues to expand in response to evolving needs, ranging from simple conversational interactions to increasingly complex tasks. As the number of human–technology interactions grows and technological capabilities advance, new challenges emerge that require effective solutions. AI is increasingly used to facilitate understanding, support decision-making processes, and enhance efficiency across various domains. This trend is particularly visible in education. According to the Digital Education Council's (2024) global survey, 86% of students use AI in their studies, with 69% employing it for information searches and 42% for grammar checking. Similarly, a survey conducted at the Hong Kong University of Science and Technology found that 96% of students use AI tools, primarily for academic purposes (Jingwei He et al., 2025).

Markowitz et al. (2024) suggested analysing the features of AI-generated texts, using writing analysis to measure how thoughts and content are transmitted. Initial findings revealed that non-academic abstracts were simpler and more approachable in terms of language complexity; full experimental detail is provided in Appendix A. Markowitz et al. (2024) observed that simpler and clearer presentation was associated with greater trust in, and understanding of, the topic, which was identified as a key advantage of AI-generated explanations. The findings were also confirmed by an independent study on the trustworthiness of AI creations conducted by Kim et al. (2024).

Feher et al. (2025) list the features of AI-generated explanations, highlighting their smooth language and high linguistic flexibility, as well as their tendency toward generalisation. What fosters trust and credibility is their high level of adaptability to the target audience through an appropriate choice of vocabulary and register. The above points lead to the conclusion that AI-generated explanations are explanations automatically produced by Large Language Models, designed to present information in a clear, coherent, and audience-oriented manner.

Credibility and Clarity

Important criteria for explanation in terms of perception are credibility and clarity, as both affect how receivers perceive the content. Credibility incorporates three elements: the sender of the message (source credibility), the channel of the message (media credibility), and the form and way in which the message is conveyed (message credibility) (Hellmueller & Trilling, 2012). This aspect constitutes the level of understanding and trust in the received content. Clarity, in turn, refers to the degree to which the message is comprehensible and unambiguous to its intended audience.

Rae (2024) conducted mixed-method experiments in which participants evaluated content created either by a computer or by humans. The findings revealed that labelling content as AI-generated reduced the creator's trustworthiness and overall positive perception to a much greater extent than in the case of human-made content. However, the trustworthiness of the content itself remained unchanged, which shows the still huge gap that exists with respect to understanding the cognitive mechanisms of trustworthiness and perception. Nazaretsky et al. (2024) explored how automated systems can effectively assess the quality of students' reasoning, detect misconceptions, and deliver personalised feedback. The study underscores the importance of explanations as indicators of conceptual understanding and highlights their potential to support more effective teaching and learning. Farrell (2025) emphasises the role of education and the theoretical background for recognising and trusting AI-generated explanations. The observation in the research is that students with advanced linguistic training are more sensitive to textual coherence, tone and discourse appropriateness, which helps them cope with explanations more critically and selectively.

Zhang et al. (2025a) found that students placed greater trust in feedback co-produced by AI together with human-generated content than in purely AI-generated material. The key determinant of trust in AI-generated content appears to be experience and critical reasoning, both of which enable deeper analysis of the resulting texts. This finding is supported by Lermann Henestrosa and Kimmerle (2024), who similarly observed that knowledge of (or even the slightest suggestion about) the authorship of an AI-generated explanation significantly affects readers' trust in its validity. The researchers suggest further research on trustworthiness and credibility, particularly by contrasting human- and AI-generated content, especially for broader audiences. An interesting finding of the paper is that AI explanations are based on the probability and frequency of specific data, which probably makes them more reliable and trustworthy than human-made content. In this respect, a comparison of groups with different national backgrounds and attitudes to AI-created content might be of great value for future research on the topic.

Research Gap and Research Questions

The study described in the following paper refers to Poland and Croatia as two countries struggling with the rapidly growing popularity of AI across many sectors, including education. To fill the existing research gap mentioned in the previous paragraph, and referring, on the one hand, to the lack of comparative studies on AI vs human explanations, and, on the other hand, to the under-researched element of the trustworthiness and reliability of AI- and human-created explanations, the following research questions were posed.

Poland and Croatia were selected on both practical grounds (the authors' institutional affiliations enabled access to comparable undergraduate cohorts) and theoretical ones (both represent underrepresented Central and Eastern European higher education systems in AI-in-education research); a detailed discussion of the national and policy context is provided in Appendix A.

The constructs in this study are positioned within a coherent conceptual framework as follows. Explanation type (AI-generated, human-created, or hybrid) is the primary independent variable; clarity, credibility, usefulness, and engagement are outcome dimensions; trust in AI-generated content functions as a mediating variable between explanation type and content preference; and student characteristics (AI use frequency, disciplinary background, gender) operate as moderating variables. This framework, reproduced in full in Appendix B, guided construct operationalisation and the sequencing of all inferential analyses.

RQ1: How do students rate the clarity, usefulness, and credibility of AI-generated, human-created, and hybrid explanations, and which type do they prefer as a learning resource?

RQ2: What factors influence students' trust in and preference for AI-generated versus human-created explanations?

RQ3: What are students' perceptions of the strengths and weaknesses of different content types, and what improvements do they recommend for AI-generated content?

Based on the above research questions, the following hypotheses were formulated:

H1: Students will rate human-created explanations as significantly clearer and more credible than AI-generated explanations.

H2: Frequency of AI use will be positively associated with higher trust in and more favourable evaluation of AI-generated content.

H3: Students' disciplinary background will significantly influence their ability to distinguish AI-generated from human-created content, with language-specialised students performing differently from STEM students.

Methodology

The study adopted a comprehensive quantitative research methodology to check the reliability and clarity of AI-generated versus human-created explanations. The data were collected through surveys distributed to participants in Poland and Croatia. The questionnaire was available via Google Forms and was completed by 298 participants between April and May, 2025.

The choice of a quasi-experimental mixed-methods design was motivated by the multi-dimensional nature of the constructs under investigation. The quantitative strand, comprising Likert-scale ratings, forced-choice source identification tasks, and preference questions, enabled statistical testing of group differences and associations across disciplines, genders, and national backgrounds. The qualitative strand, comprising open-ended responses analysed thematically, allowed students to articulate their evaluative reasoning in their own words, capturing nuance that closed-scale items cannot fully represent. These two strands were integrated at the interpretive level: quantitative results identified patterns and tested the study's hypotheses, while qualitative data explained the mechanisms and contextual factors underlying those patterns. The quasi-experimental character of the design derives from the use of controlled stimulus texts (matched pairs of AI-generated and human-created explanations on the same topic) presented to participants under standardised conditions, without random assignment.

Each construct was operationalised through specific survey items; full definitions and item wording are provided in Appendix B (Table B.1). All Likert items used a 5-point response scale (1 = strongly disagree / not at all, 5 = strongly agree / very much).

Item wording was developed based on established frameworks for assessing credibility, clarity, and trust in educational contexts (Hellmueller & Trilling, 2012; Lermann Henestrosa & Kimmerle, 2024) and adapted to the specific task of evaluating short explanatory texts. Single-item operationalisation was adopted as a pragmatic compromise given survey length constraints, with face validity established through the pilot phase ($N = 15$). The limitations of this approach regarding construct validity and internal consistency are explicitly acknowledged in the Limitations section.

The primary aim of the survey was to assess university students' ability to identify the source of explanatory texts (AI-generated vs. human-created) and to evaluate those texts along dimensions of clarity, credibility, usefulness, and engagement. The target group consisted of undergraduate students enrolled in higher education institutions in Croatia and Poland, selected on the basis of availability and voluntary participation (convenience sampling). This non-probability sampling approach was chosen to facilitate data collection within the institutional constraints of both universities; however, it should be noted that this limits the generalizability of findings to broader student populations. The final sample comprised 298 respondents (198 from Poland, 100 from Croatia). Prior to the survey, a pilot test was conducted with a small group of students ($N = 15$) to evaluate clarity of instructions and item comprehension; no significant issues were identified. Participation was entirely voluntary and anonymous, and no minimum sample size was predefined; the achieved $N = 298$ is considered adequate for the non-parametric statistical tests applied (Wilcoxon signed-rank, Mann-Whitney U, Kruskal-Wallis, chi-square).

Participants

Participation was voluntary, and all participants provided informed consent after being informed about anonymity and data-handling procedures. The analysis was conducted on 298 people, including 198 from Poland and 100 from Croatia. In the survey, 135 male participants took part, 152 female, and 11 who preferred not to state their gender. The vast majority of the participants belonged to the age group 18–21, 71 to the group 22–24 and 13 to the group 25 and older. 90.9% ($N = 271$) of the participants were at the bachelor level. 90.3% ($N = 269$) of the participants were enrolled at a public university. 38% of the survey participants study Computer Science, 37% Modern Languages, 17% Applied Linguistics, 6% Business, 1% Management and Medicine and 0% Psychology.

Data Analysis

Quantitative responses were analysed using descriptive statistics and non-parametric inferential tests (Wilcoxon signed-rank, Mann-Whitney U, Kruskal-Wallis and chi-square tests), appropriate for ordinal data and subgroup comparisons. Open-ended responses were subjected to thematic analysis, which identified recurring patterns in students' evaluative reasoning and perceptions of AI versus human authorship.

Figure 2 presents the study's analytical framework, explicitly mapping each research question and hypothesis to its corresponding construct, statistical test, and variable(s). This framework guided the selection and sequencing of all inferential analyses reported in the Results section. The analytical framework is reproduced in full in Appendix B.

Results

This section presents the study's findings, structured around five analytical themes that collectively address the three formal research questions stated above.

Terminology note: 'credibility' refers to text-level perceived accuracy; 'trust' to the broader disposition toward AI as a source type; full definitions and effect sizes for all significant tests are provided in Appendix D.

The survey results revealed that 88.3% of the student sample actively integrates AI technologies into their everyday activities, driven primarily by efficiency, convenience, and their role as study aids. The remaining 11.7% cited a lack of personal need, concerns about misinformation, or a preference for independent inquiry as reasons for non-use. A full categorisation of AI usage patterns and reasons for avoidance is provided in Appendix D.

A Wilcoxon signed-rank test ($N = 298$) was conducted to compare students' perceived difficulty in understanding explanations created by humans and generated by AI. The test yielded a significant result ($Z = –3.263$, $p = .001$) with 120 students (40.3%) finding AI-generated material easier to understand, 73 (24.5%) reporting greater difficulty, and 105 (35.2%) noticing no difference.

A subsequent Mann-Whitney U test assessed differences between daily AI users ($N = 263$) and non-users ($N = 35$) in perceived comprehension of AI-generated content. Daily students reported significantly lower difficulty ($U = 3035.500$, $z = –3.492$, $p = .000$), suggesting that frequent exposure may enhance students' ability to comprehend AI-generated content, possibly due to increased familiarity with the structure, language patterns and limitations of such content. Similarly, daily students exhibited greater trust in AI-generated content ($U = 1784.500$, $z = –6.287$, $p = .000$).

Cross-national comparisons (Mann-Whitney U) revealed no significant differences between Polish and Croatian students in either perceived difficulty of AI-generated explanation ($U = 9259.500$, $z = 0.973$, $p = .330$) or trust in AI as a source of academic knowledge ($U = 9102.00$, $z = -1.214$, $p = .225$); full cross-national analysis is provided in Appendix D.

A within-group Kruskal-Wallis test revealed significant discipline-based variation among the Polish students ($X^2(6, N = 198) = 12.152$, $p = .002$). Post-hoc pairwise comparisons (with Bonferroni correction) showed that students in Philology and Applied Linguistics found AI-generated content significantly more difficult than students in other fields ($p = .015$ and $p = .016$, respectively).

Discussion

This study examined university students' perceptions of AI-generated versus human-created explanations by exploring three central research questions: how students assess and prefer different content types; what factors influence their trust and preferences; and how they perceive the strengths and weaknesses of each, along with suggestions for improvement. The findings provide valuable insights into how students engage with AI tools in educational contexts and offer implications for both AI literacy development and the design of AI-supported instructional content.

The first research question addressed how students rate the clarity, usefulness and credibility of explanations depending on their source and which type they ultimately prefer. The analysis revealed that human-created content was consistently rated as clearer, more useful and more credible than AI-generated content, particularly in logic-related tasks. Students appreciated the accessible language, personal tone and practical examples found in human-authored texts, which facilitated comprehension and engagement. However, AI-generated content was valued for its depth, precision and structured presentation, attributes that were particularly appreciated in more technical subjects such as neurobiology. One possible explanation is that AI-generated content often uses simpler, more structured and more concise language, which may reduce cognitive load and make complex concepts more accessible. In contrast, human-created content may include more nuanced or context-dependent expressions, which, while potentially richer in meaning, can be harder to process for some learners. This pattern aligns with the findings of Nazaretsky et al. (2024). Despite these perceived strengths, the AI-generated explanations were sometimes seen as overly formal or linguistically rigid, especially by students with advanced language training. Interestingly, across both domains and regardless of national background, students most frequently expressed a preference for hybrid content - explanations that combine human contextual clarity with AI-generated precision. This suggests a growing awareness of the complementary strengths of both content types and a preference for a balanced integration that maximises both clarity and informational value.

The second research question focused on the factors shaping students' trust in and preferences for AI-generated versus human-created content. Several key variables emerged from the analysis. First, the frequency of AI use was strongly associated with more favourable evaluations of AI-generated content.

Students who reported daily or frequent use of AI tools found such content easier to understand and more trustworthy, echoing findings by Zhang et al. (2025b) and Nazaretsky et al. (2024) on the role of exposure in reducing cognitive load and increasing tool trust, i.e. the more students engage with AI, the more adept they become at parsing its linguistic structure and content delivery.

This indicates that regular exposure to AI tools may contribute to a greater sense of reliability and confidence in AI as a source of academic information. The findings are consistent with prior research (Zhang et al., 2025a), which suggests that frequency of AI use is positively associated with both perceived usefulness and trust in AI-generated educational materials. Second, disciplinary background emerged as a significant factor influencing both perception and recognition. Philology and Applied Linguistics students rated AI-generated content as more difficult to understand than students in other fields. Our result aligns with Farrell's (2025) research. These findings suggest that students with strong backgrounds in language and textual analysis may be more sensitive to the limitations or unnatural features of AI-generated texts, such as a lack of nuance, coherence or pragmatic appropriateness. Their academic background likely enables them to detect subtle inconsistencies or unnatural phrasing that may be overlooked by peers from other disciplines. Our result aligns with Farrell's (2025) research, which highlights that language-specialised students engage more critically with machine-produced texts. These results imply that while most students may perceive AI-generated explanations as accessible or even preferable, those with deep linguistic expertise may remain cautious. This underscores the importance of enhancing the communicative clarity and authenticity of AI-generated educational materials, especially when used in language-focused academic contexts.

In contrast, Computer Science students not only evaluated AI explanations more positively but also outperformed other groups in source identification tasks. This may reflect their greater exposure to AI tools and conceptual familiarity with algorithmic logic, enabling more accurate evaluations and increased receptivity toward machine-generated content. These findings highlight the influence of academic training on students' critical engagement with AI-produced content. Third, gender also influenced perception and comprehension, although to a lesser extent. Female students reported lower difficulty in understanding AI-generated materials, a finding that resonates with studies identifying greater comfort and trust in digital tools among women in academic settings (Zhang et al., 2025a). However, no significant gender differences were found in trust ratings overall. Interestingly, male students perceived the human-created neurobiology explanation as significantly easier to understand and more useful, indicating possible content-specific confidence or greater prior knowledge in technical subjects.

Taken together, the findings of the present study both confirm and extend existing research in important ways. The preference for hybrid content aligns with Shankar et al. (2024), who argue that incorporating human oversight into AI-generated material enhances its pedagogical credibility. Similarly, the strong association between frequent AI use and more favourable content evaluations resonates with Zhang et al. (2025a), who found that students accustomed to AI tools perceive their output as more objective and useful, a pattern consistent with technology acceptance theory, which posits that familiarity reduces perceived risk. However, our findings diverge from Huschens et al. (2023), who reported higher clarity ratings for AI-generated texts; in our study, human-created content was consistently rated as clearer and more engaging. This discrepancy may reflect differences in participant profiles or task design, and warrants replication across broader populations. Furthermore, while prior research (e.g., Clark et al., 2021; Waltzer et al., 2024) has reported relatively low accuracy in identifying AI-generated text, our students achieved markedly higher accuracy (86% in the logic domain), suggesting that the level of AI familiarity in a given cohort may be a critical moderating variable – one that the existing literature has not yet fully addressed. Future research should examine whether AI literacy programs can be designed to capitalise on this emerging competence, and whether similar patterns hold across different text genres, languages, and educational systems. Longitudinal designs are particularly needed to track how student perceptions and detection abilities evolve as AI tools themselves continue to develop.

A further dimension that warrants reflection concerns the potential role of cultural differences between the Polish and Croatian respondents. Although the cross-national statistical comparisons yielded no significant differences, a null result does not preclude the possibility that cultural factors shape students' engagement with AI-generated content in ways not captured by the measures employed. While the two countries share a Central and Eastern European educational heritage, they differ in language family, digital infrastructure, and institutional AI policy frameworks, all of which may influence AI-related attitudes through channels not covered by this study. Future research should incorporate culturally sensitive instruments and qualitative follow-up to examine whether the apparent equivalence of responses masks more nuanced cross-cultural differences in how students interpret and trust AI-generated academic content.

Limitations

While the results are promising, the study is not without limitations. The sample was limited to Croatian and Polish university students, which may constrain the generalizability of the findings to other cultural and educational contexts. The content used in the source recognition task was deliberately short and focused, which may have facilitated higher identification rates than would occur with more complex or ambiguous material. Self-reported data may also be subject to bias or variability in interpretation. Future research should extend these findings through longitudinal studies, cross-cultural comparisons and the inclusion of additional academic disciplines to better understand how students' perceptions of and interactions with AI evolve over time.

Several additional limitations merit acknowledgement. The sample is uneven across disciplines (Computer Science and Modern Languages together account for 75% of participants), so comparisons involving smaller subgroups should be interpreted with caution. The high rate of reported AI use (88.3%) may reflect self-selection bias, limiting generalizability to less AI-literate populations. The exceptionally high source-identification accuracy (86%) is partially attributable to the use of short, topically constrained texts and a binary forced-choice format, which may have inflated accuracy relative to more naturalistic conditions. All constructs were measured using single self-report items rather than validated multi-item scales, which limits construct validity; future studies should develop and validate multi-item instruments for this context. Finally, the study does not account for participants' prior topic knowledge, which may have influenced source recognition independently of disciplinary background.

Conclusion

This study explored university students' perceptions and evaluations of AI-generated versus human-created explanation, their ability to distinguish between the two, and how factors such as discipline, gender, and frequency of AI use influence these perceptions. The results revealed a high rate of AI integration in the students' academic and personal lives, coupled with a predominantly pragmatic attitude toward its use. While students recognised the efficiency and depth of AI-generated explanations, they consistently valued the clarity, relatability and authenticity of human-created content. A clear preference emerged for hybrid explanations that combined the strengths of both sources.

One of the most significant findings was the students' high accuracy in identifying the source of explanations, challenging the common assumption that AI-generated texts are difficult to detect. This accuracy appeared to be closely linked to their familiarity with AI tools, critical evaluation skills and perceptions of clarity, usefulness and trustworthiness. However, the ability to distinguish between AI-generated and human-created content was shown to be domain-sensitive, with higher performance in familiar, logic-based tasks and reduced accuracy in more complex, specialised fields such as neurobiology.

With respect to trust (a central concern of the study), the findings present a nuanced picture. Students did not unconditionally trust AI-generated content; rather, trust was conditional and context-dependent. It was strongest among frequent AI users, in technically complex domains where AI explanations were perceived as authoritative (e.g., neurobiology), and when AI-generated content was framed as reviewed or validated by a human expert (as supported by 167 respondents). Trust was lower for unfamiliar or evaluatively demanding content and among students with strong linguistic training. These patterns suggest that trust in AI-generated explanations is not simply a function of technological acceptance, but is mediated by domain familiarity, AI literacy, and the perceived presence of human oversight. This has direct implications for educational practice: designing AI-supported learning environments that include transparent human curation or validation mechanisms may be the most effective strategy for building appropriate, calibrated trust in AI-generated educational content.

While the results offer promising insights into students' emerging competencies in navigating AI-supported learning environments, the study is limited by its focus on students from only two national contexts and its reliance on short-form texts. Future research should expand to include a broader range of disciplines, educational settings and text genres. Longitudinal and intervention-based studies would also help to clarify how evaluative skills and trust in AI evolve over time and with targeted training.

The appendices are available in the online version of the journal.

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the ethical standards of the authors' higher education institutions. All participants were informed about the purpose of the research, their rights as participants, and data protection measures. Informed consent was obtained from all individual participants involved in the study.

Availability of data and material

The datasets generated and analysed during the current study are available from the corresponding author upon reasonable request. Any data shared will be anonymised to ensure participant confidentiality.

Acknowledgements

The authors wish to thank all student participants for their time and valuable insights.

References

Berber Sardinha, T. (2024). AI-generated vs human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics, 4*(1), 100083. https://doi.org/10.1016/j.acorp.2023.100083

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*(43). https://doi.org/10.1186/s41239-023-00411-8

Clark, E., August, T., Serrano, S., Haduong, N., Gururangan, S., & Smith, N. A. (2021). All that's 'human' is not gold: evaluating human evaluation of generated text. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, vol. 1 (pp. 7282–7296). https://doi.org/10.18653/v1/2021.acl-long.565

Denny, P., Khosravi, H., Hellas, A., Leinonen, J., & Sarsa, S. (2023). *Can we trust AI-generated educational content? Comparative analysis of human and AI-generated learning resources*. arXiv. https://doi.org/10.48550/arXiv.2306.10509

Digital Education Council. (2024). *Global AI Student Survey*. https://www.digitaleducationcouncil.com/resource-library-items/digital-education-council-global-ai-student-survey-2024

Farrell, M. (2025). Can postgraduate translation students identify machine-generated text? In *Proceedings of Machine Translation Summit XX,* vol. 1 (pp. 432–441). https://aclanthology.org/2025.mtsummit-1.34.pdf

Federer, M. R., Nehm, R. H., Opfer, J. E., & Pearl, D. (2015). Using a constructed-response instrument to explore the effects of item position and item features on the assessment of students' written scientific explanations. *Research in Science Education, 45*, 527–553. https://doi.org/10.1007/s11165-014-9435-9

Feher, D., Khered, A., Zhang, H., Batista-Navarro, R., & Schlegel, V. (2025). Learning to generate and evaluate fact-checking explanations with transformers. *Engineering Applications of Artificial Intelligence, 139* (Part A), 109492. https://doi.org/10.1016/j.engappai.2024.109492

França, C. (2023). *AI empowering research: 10 ways how science can benefit from AI*. arXiv. https://arxiv.org/abs/2307.10265

Hellmueller, L., & Trilling, D. (2012), The credibility of credibility measures: A meta-analysis in leading communication journals, 1951 to 2011. In *WAPOR Hong Kong 2012*: paper presentation World Association for Public Opinion Research/Public Opinion Programme, The University of Hong Kong.

Huschens, M., Briesch, M., Sobania, D., & Rothlauf, F. (2023). *Do you trust ChatGPT? Perceived credibility of human and AI-generated content*. arXiv. https://doi.org/10.48550/arXiv.2309.02524

Jingwei He, A., Zhang, Z., Anand, P., & McMinn, S. (2025). Embracing generative artificial intelligence tools in higher education: a survey study at the Hong Kong University of Science and Technology. *Journal of Asian Public Policy*, *18*(2), 352–376. https://doi.org/10.1080/17516234.2024.2447195

Keil, F. C., & Wilson, R. A. (2000). Explaining explanation. In F. C. Keil, & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 1–18). MIT Press. https://doi.org/10.7551/mitpress/2930.003.0003

Josephson, J. R., & Josephson, S. G. (1996). *Abductive inference: Computation, philosophy, technology*. Cambridge University Press. https://doi.org/10.1017/CBO9780511530128

Kim, J. H., Kim, J., Kim, C., & Kim, S. (2023). Do you trust ChatGPTs? Effects of the ethical and quality issues of generative AI on travel decisions. *Journal of Travel & Tourism Marketing, 40*(9), 779–801. https://doi.org/10.1080/10548408.2023.2293006

Kim J., Maathuis, H., & Sent, D. (2024). Human-centered evaluation of explainable AI applications: a systematic review. *Frontiers in Artificial Intelligence, 7*. https://doi.org/10.3389/frai.2024.1456486

Lee, S., & Song, K.-S. (2024). Teachers' and students' perceptions of AI-generated concept explanations: Implications for integrating generative AI in computer science education. *Computers and Education: Artificial Intelligence, 7*, 100283. https://doi.org/10.1016/j.caeai.2024.100283

Leinhardt, G., & Steele, M. D. (2005). Seeing the complexity of standing to the side: Instructional dialogues. *Cognition and Instruction, 23*(1), 87–163. https://doi.org/10.1207/s1532690xci2301_4

Lermann Henestrosa, A., & Kimmerle, J. (2024). The effects of assumed AI vs. human authorship on the perception of a GPT-generated text. *Journalism and Media, 5*(3), 1085–1097. https://doi.org/10.3390/journalmedia5030069

Lombrozo, T., & Liquin, E. G. (2023). Explanation is effective because it is selective. *Current Directions in Psychological Science, 32*(3), 212–219. https://doi.org/10.1177/09637214231156106

Markowitz, D. M., Hancock, J. T., & Bailenson, J. N. (2024). Linguistic markers of inherently false AI communication and intentionally false human communication: Evidence from hotel reviews. *Journal of Language and Social Psychology, 43*(1), 63–82. https://doi.org/10.1177/0261927X231200201

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

Moore, B. A., & Wright, J. (2023). Constructing written scientific explanations: a conceptual analysis supporting diverse and exceptional middle- and high-school students in developing science disciplinary literacy. *Frontiers in Education, 8,* 1305464. https://doi.org/10.3389/feduc.2023.1305464

Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T. (2024). AI or human? Evaluating student feedback perceptions in higher education. In R. Ferreira Mello, N. Rummel, I. Jivet, G. Pishtari, & J. A. Ruipérez Valiente (Eds.), *19th European Conference on Technology Enhanced Learning, EC-TEL 2024* (pp. 284–298). Springer. https://doi.org/10.1007/978-3-031-72315-5_20

Pitts, G., Marcus, V., & Motamedi, S. (2025). *Student perspectives on the benefits and risks of AI in education*. arXiv. https://arxiv.org/abs/2505.02198

Poggiolessi, F. (2024). (Conceptual) explanations in logic. *Journal of Logic and Computation, 35*(4). https://doi.org/10.1093/logcom/exae064

Rae, I. (2024). The effects of perceived AI use on content perceptions. In *CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3613904.3642076

Rago, A., Palfi, B., Sukpanichnant, P., Nabli, H., Vivek, K., Kostopoulou, O., Kinross, J., & Toni, F. (2024). *Exploring the effect of explanation content and format on user comprehension and trust in healthcare*. arXiv. https://doi.org/10.48550/arXiv.2408.17401

Shankar, M., Hazfiarini, A., Zahroh, R. I., Vogel, J. P., McDougall, A. R. A., Condron, P., Goudar, S. S., Pujar, Y. V., Somannavar, M. S., Charantimath, U., Ammerdorffer, A., Rushwan, S., Gülmezoglu, A. M., & Bohren, M. A. (2024). Factors influencing the participation of pregnant and lactating women in clinical trials: A mixed-methods systematic review. *PLoS Medicine, 21*(5), e1004405. https://doi.org/10.1371/journal.pmed.1004405

Tan, C. (2022). On the diversity and limits of human explanations. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 2177–2183). https://doi.org/10.18653/v1/2022.naacl-main.158

Tu, J., Hadan, H., Wang, D. M., Sgandurra, S. A., Mogavi, R. H., & Nacke, L. E. (2024). *Augmenting the author: Exploring the potential of AI collaboration in academic writing.* arXiv. https://doi.org/10.48550/arXiv.2404.16071

Waltzer, T., Pilegard, C., & Heyman, G. D. (2024). Can you spot the bot? Identifying AI-generated writing in college essays. *International Journal for Educational Integrity, 20*(11). https://doi.org/10.1007/s40979-024-00158-3

Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford University Press.

Zhang, A., Gao, Y., Suraworachet, W., Nazaretsky, T., & Cukurova, M. (2025a). *Evaluating trust in AI, Human, and* *co-produced feedback among undergraduate students*. arXiv. https://doi.org/10.48550/arXiv.2504.10961

Zhang, T., Zhang, M., Low, W. Y., Yang, X. J., & Li, B. A. (2025b). Conversational explanations: Discussing explainable AI with non-AI experts. In *IUI '25: Proceedings of the 30th International Conference on Intelligent User Interfaces* (pp. 409–424). https://doi.org/10.1145/3708359.3712143

Dragana Božić Lenard is an Associate Professor at the Faculty of Electrical Engineering, Computer Science and Information Technology (FERIT) at Josip Juraj Strossmayer University of Osijek, Croatia, where she teaches English for Specific Purposes courses. She holds a doctorate degree from the Faculty of Humanities and Social Sciences, Osijek and has authored over 30 peer-reviewed publications, with research interests spanning sociolinguistics, English for Specific Purposes (ESP), teaching methodology, grammar and vocabulary acquisition, students' learning preferences and pedagogical content knowledge. Her scholarly work explores how teachers and learners develop linguistic and cultural identities through foreign language instruction, with a particular focus on the contexts of ESP and English as a Foreign Language (EFL). She is the author of two university coursebooks in English for computer science. An active member of the international academic community, Dr. Božić Lenard serves as lead editor of the ERL Journal and co-organiser of the ERL Online Sessions within the Educational Role of Language Association (ERLA). She is a firm advocate of the principle that every teacher is a language teacher and works to strengthen the connection between language and educational sciences across disciplines.

Joanna Kic-Drgas is an Associate Professor affiliated with the Institute of Applied Linguistics at Adam Mickiewicz University in Poznań. She completed her doctorate in 2012 at the same institution, specialising in foreign language didactics for older adults, and has been working there as a research fellow since 2014. She has a decade of experience in teaching languages for specific purposes (LSP) and has authored and co-authored numerous works focused on teaching methodology, intercultural communication, and LSP/ESP. Her research interests span applied linguistics, foreign language geragogy (language learning in older age), teacher education, and the integration of technology in language teaching. Her more recent scholarly work examines the role of artificial intelligence in language education, including research on AI-driven chatbots that foster collaborative learning among pre-service EFL teachers. She has also contributed to international research projects involving augmented reality and mobile learning tools for language acquisition.