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

Appendices

Appendix A: Extended Literature Review Passages

A.1 The Nature of Explanation: Extended Discussion

The complex nature of explanation suggests that it is deeply rooted in the understanding of a given problem, which in turn depends on previously acquired knowledge. First, explanations do not merely reflect an individual's way of thinking or approach to solving a problem; rather, they emerge as a response to a knowledge gap. Second, explanations seek to answer the question of why. Through reasoning, mathematical proofs, and the use of generalizations, they identify key points of reference and, on this basis, contribute to shaping an individual's understanding of reality. Explanations function as interpretations—that is, individual constructions of specific phenomena—and are therefore developed within the cognitive framework of the speaker.

Furthermore, explaining involves exploring new and previously uncharted areas of knowledge in order to better understand their practical significance. Explanations can also serve as a foundation for generating new explanations, encouraging the adoption of alternative perspectives and deeper insights. Finally, explanations play a crucial communicative role, facilitating the exchange of knowledge and understanding among people from different professional, age, and interest groups.

A.2 Rago et al. (2024): Experimental Detail

An experiment conducted by Rago et al. (2024) contributed to confirming the correlation between the format and clarity of an explanation and the level of trust in it. The research group consisted of both medical students and the general sample (with limited medical experience, representing stakeholders in the form of patients). Both groups were exposed to explanations in the form of texts and charts. In the experiment the following aspects were measured: objective understanding (with questions related to reading comprehension), subjective understanding (individual perception of how understandable a given text is for the participant) and trust in the explanations obtained. Text-based explanations were rated as more understandable and more trustworthy.

A.3 Markowitz et al. (2024): Experimental Detail

Markowitz et al. (2024) suggested analyzing the features of AI generated texts, using writing analysis to measure how thoughts and content are transmitted. In the first phase of the experiment, the researchers used abstracts from both scientific and popular journals. Initial findings after this stage revealed that the non-academic abstracts were simpler and more approachable in terms of language complexity. Based on the abstracts and the initial findings, "significance statements" were generated by AI with a level of simplicity two levels lower than the language analysis of the original texts showed.

In the next stage, the participants of the study read the texts generated by both AI and humans, and assessed the trustworthiness and intelligence of the writers, and their reliability. In the case of the AI generated texts, the reliability and trustworthiness of the text was higher, but the intelligence of the authors was assessed as lower than in the case of the human generated texts. The results of the study showed that the simplifications affected the clarity and understanding of the texts positively, yet they affected the perception of competence negatively. The texts generated by AI contributed to higher engagement with the research communication. The paradox observed by the researcher is the simple and clear presentation of thoughts – the greater trust and understanding of the presented problem, which was the main advantage of the explanations created by AI (Markowitz et al., 2024). The findings were also confirmed by an independent study on the trustworthiness of AI creations conducted by Kim et al. (2024).

A.4 Research Context: Poland and Croatia

The comparison between Poland and Croatia is justified by their similar post-socialist backgrounds, which continue to influence the structure and functioning of their educational systems. Both countries remain relatively underrepresented in the growing body of research on AI literacy. Therefore, conducting a joint study may not only increase the visibility of Poland and Croatia within the international scientific community but also complement existing research carried out in other national contexts. As member states of the European Union, both countries shape their AI-related policies within the framework of shared supranational regulations. However, the implementation and interpretation of these regulations may differ at the national level. Of particular research interest is the extent to which these differences influence students' levels of digital literacy and their sensitivity to AI-generated texts. The empirical evidence gathered through this study may provide valuable insights for the development of educational curricula that effectively integrate AI tools while fostering critical awareness and responsible AI use.

Appendix B: Analytical Framework and Survey Instrument Design

B.1 Survey Construct Operationalization

Each construct was operationalized through specific survey items. Clarity was measured by a rating item asking participants to assess how easy it was to understand the explanation. Usefulness was assessed by an item asking whether the explanation would help them learn or understand the topic. Credibility was assessed by an item asking how reliable and accurate they considered the information in the explanation. Engagement was assessed by asking how interesting or engaging they found the explanation. Trust was operationalized as a broader attitudinal item asking about confidence in AI as a source of academic knowledge. Content preference was measured by a forced-choice item asking which of the three explanation types (human-created, AI-generated, or hybrid) they would prefer as a learning resource. Credibility and trust are treated as distinct: credibility is assessed at the level of a specific text, while trust reflects a general orientation toward AI-generated content as a source type. All Likert items used a 5-point response scale (1 = strongly disagree / not at all, 5 = strongly agree / very much).

Table B.1

Survey Constructs, item Wording, and Response Scales

Construct	Survey Item	Response Scale
Clarity	How easy was it to understand this explanation?	1–5 (1 = not at all, 5 = very much)
Usefulness	Would this explanation help you learn or understand the topic?	1–5 (1 = not at all, 5 = very much)
Credibility	How reliable and accurate do you consider the information in this explanation?	1–5 (1 = not at all, 5 = very much)
Engagement	How interesting or engaging did you find this explanation?	1–5 (1 = not at all, 5 = very much)
Trust (general)	How confident are you in AI as a source of academic knowledge?	1–5 (1 = not at all confident, 5 = very confident)
Content preference	Which of the three explanation types (human-created, AI-generated, or hybrid) would you prefer as a learning resource?	Forced choice: Human / AI / Hybrid

B.2 Analytical Framework

Image 2 presents the analytical framework of the study, explicitly mapping each research question and hypothesis to its corresponding construct, statistical test, and the variable(s) involved. This framework guided the selection and sequencing of all inferential analyses reported in the Results section.

Table B.2

Analytical Framework: Research Questions, Hypotheses, Constructs, and Statistical Tests

Hypothesis	Construct measured	Statistical test	Independent variable	Dependent variable
H1: Human > AI on clarity and credibility	Clarity, usefulness, credibility, engagement, content preference	Wilcoxon signed-rank test (paired: same participants rate both text types); chi-square (preference distribution)	Content type (AI-generated vs. human-created)	Ratings of clarity, usefulness, credibility, engagement; preference choice
H2: AI-use frequency positively associated with trust in AI content	Trust in AI-generated content; perceived comprehension of AI content	Mann-Whitney U test (daily vs. non-daily AI users); national comparison (Polish vs. Croatian students)	AI-use frequency; national background; gender	Trust rating; perceived difficulty of AI content
H3: Disciplinary background influences source-detection accuracy	Source-identification accuracy; perceived comprehension difficulty across disciplines	Kruskal-Wallis test (discipline groups); Mann-Whitney U (pairwise post-hoc with Bonferroni correction)	Academic discipline	Perceived difficulty of AI content; source-identification accuracy rate
H4: Exploratory	Perceived strengths and weaknesses of each content type; improvement suggestions	Thematic analysis of open-ended responses	N/A (qualitative)	Thematic categories from open-ended responses

B.3 Survey Access

The questionnaire was available via Google Forms (https://docs.google.com/spreadsheets/d/1qiOie9Ny6XJjw8euaxBjeoY46p9wkhRyxnsiAnwiRw/edit?usp=sharing) and was completed by 298 participants between April and May, 2025. The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request. Any data shared will be anonymized to ensure participant confidentiality.

Appendix C: Extended Results Analysis

C.1 Source Identification Accuracy: Logic Task

Contrary to findings in the existing literature, our study revealed that 86% of the students correctly distinguished between AI-generated and human-created content. Accuracy was measured using a forced-choice task in which each student was presented with two explanatory texts - one written by a subject-matter expert and the other generated by an AI system - both addressing the same topic in logic. The students were asked to identify the source of each explanation with no time constraints imposed. This notably high accuracy challenges the prevailing assumption that such distinctions are inherently difficult. In Waltzer et al.'s (2024) study, college instructors identified AI-generated essays correctly only about 70% of the time, while students performed worse, around 60% accuracy. Similarly, linguistics experts were able to detect AI-generated text just 38.9% of the time, often with inconsistent reasoning (Clark et al., 2021). These results have led many scholars to argue that traditional detection methods are insufficient and that improving AI literacy and rethinking assessment practices are crucial for maintaining academic integrity in the age of generative AI. That our students achieved 86% accuracy is particularly striking given these precedents. This result suggests that under certain conditions, such as when content is concise, focused and topically constrained, students may be far more capable of distinguishing between AI and human authorship than previously assumed. A potential explanation lies in our students' high familiarity with AI tools: 88.3% reported daily or frequent AI use, which may have enhanced their sensitivity to the linguistic and stylistic patterns typical of AI-generated content. Additionally, our use of brief, targeted explanatory texts rather than full-length essays may have accentuated those linguistic cues (e.g. tone, structure, precision) that students have learned to associate with AI writing. These findings highlight the need to reassess assumptions about AI detectability and point to the potential value of integrating structured exposure to AI writing into curricula to strengthen students' evaluative competence.

C.2 Quality Evaluation and Source Recognition: Logic Task

The results of the Mann-Whitney U tests indicated statistically significant differences between students who correctly identified the source of the content (i.e. human-created or AI-generated) and those who failed to do so. Students who rated the explanations as easier to understand ($U = 3551.000$, $p = .016$), more useful ($U = 4020.000$, $p = .048$), more reliable ($U = 3270.000$, $p = .001$), and accompanied by more relevant examples ($U = 3928.000$, $p = .017$) were significantly more likely to accurately determine the origin of the content. These findings suggest that positive evaluations of explanatory quality, including clarity, usefulness, relevance of examples and perceived reliability, are strongly associated with improved accuracy in source recognition. This aligns with the previous research indicating that students' trust in content and engagement with AI-generated materials are closely tied to their perceptions of the content's comprehensibility and pedagogical value (Denny et al., 2023; Lee & Song, 2024; Zhang et al., 2025). Such insights underscore the importance of designing AI-generated educational materials that are not only factually accurate but also pedagogically effective in order to foster critical awareness and informed content evaluation.

The Wilcoxon signed-rank test revealed statistically significant differences in the students' evaluations of clarity, usefulness, credibility and engagement between human-created and AI-generated content. Human-created content was perceived as significantly clearer than AI-generated content ($Z = -5.697$, $p = .000$), a finding that contrasts with Huschens et al. (2023), who reported higher clarity ratings for AI-generated texts. Similarly, the students rated human-created content as significantly more useful ($Z = -5.230$, $p = .000$), more credible ($Z = -2.786$,

p = .005) and more engaging (Z = -7.346, p = .000). These findings partially diverge from Zhang et al. (2025b), who found that undergraduate students rated AI-human co-produced feedback as more useful and objective than fully human-created responses. These results indicate that students may associate human writing with a more authentic, nuanced or personally resonant communication style, leading to higher ratings across subjective dimensions. Furthermore, in line with our results, Huschen et al. (2023) explored perceptions of credibility and found that while AI-generated content was often rated as competent and articulate, it did not necessarily translate into higher credibility compared to human-created content. Trustworthiness remained closely linked to perceived human authorship, particularly in contexts requiring subjective judgment or emotional resonance. This trend may reflect a residual bias toward human writing, or a sensitivity to subtle linguistic cues, such as tone, structure, and natural variation, that distinguish human discourse from AI-generated output. It also implies that, despite the increasing fluency of AI tools, human-created content may still be perceived as more authentic and contextually appropriate in educational settings.

C.3 Content Type Preferences: Logic Task

To determine whether the students exhibited a significant preference among the three explanation types (human-created, AI-generated and hybrid), a chi-square test of independence was conducted. Overall, 38.3% (N = 114) preferred human-created explanations, 48% (N = 143) chose hybrid and only 13.8% (N = 41) favored purely AI-generated content. The results revealed a statistically significant difference in the distribution of responses ($\chi^2(6, N = 274) = 20.64$, p = .001). A further analysis based on the students' field of study indicated that the students studying Computer Science, Philology and Applied Linguistics showed a clear preference for human-created (38.9%, 33.6%, 47%, respectively) and hybrid explanations (45.1%, 51.8%, 49%, respectively) over AI-generated content (15.9%, 14.5%, 3.9%, respectively). This indicates that students' academic backgrounds may shape their evaluation criteria when assessing content quality. For instance, students studying Philology and Applied Linguistics may be more attuned to nuances in language, tone and structure, which could lead to a stronger preference for human and hybrid explanations. Conversely, students studying Computer Science may value a balance between coherence and technical precision, favoring hybrid explanations that blend human fluency with AI efficiency. It should be noted, however, that discipline-based comparisons are limited to the three largest subgroups (Computer Science, Modern Languages, Applied Linguistics); the remaining disciplines (Business, Management, Psychology) had very small cell sizes and are therefore excluded from inferential comparisons in this section.

The students consistently indicated a preference for human-created explanations due to their clarity, simplicity and accessibility. The language used in human-created content was perceived as more straightforward and less formal, making it easier to comprehend across varying levels of language proficiency. Additionally, the students valued the personal and conversational tone of human explanations, which enhanced relatability and engagement. The inclusion of concrete examples and analogies in human-generated content was frequently highlighted as a key factor in facilitating understanding and retention. Such illustrative elements helped to contextualize abstract concepts, making them more tangible and memorable.

Furthermore, tailoring explanations to accommodate varying levels of user expertise is recommended. This could be achieved by offering layered content, ranging from simplified summaries to more detailed academic elaborations. As Tu et al. (2024) demonstrate, prompt engineering and adaptive AI responses are crucial in customizing content for different user needs and maintaining educational value. Incorporating human oversight and expert validation in the AI content creation process further enhances credibility, reliability, and pedagogical effectiveness (Shankar et al., 2024).

Many respondents ($N = 167$) express a cautious but generally positive attitude toward trusting AI-generated explanations when they have been reviewed or verified by teachers. The consensus is that while AI can provide rapid and broad access to information, it can be inaccurate, outdated or misleading. Educators, on the other hand, are valued for their specialized knowledge, experience and ability to critically assess and clarify information, which increases the reliability and trustworthiness of explanations. Many believe that an educator's review adds an essential human element, empathy, contextual understanding and pedagogical expertise that AI currently lacks. This combination is seen as beneficial in reducing errors and ensuring clarity, especially for learners. However, it is also recognized that educators are not infallible, and independent verification remains important. A minority of respondents ($N = 24$) remain skeptical about AI's role in education, citing concerns about misinformation and over-reliance on technology.

C.4 Source Identification Accuracy: Neurobiology Task

Another excerpt presented to the students for source identification was related to neurobiology. The success rate in identifying whether the content was AI-generated or human-created was notably lower (62.8%) compared to the first task, which focused on a logic-related topic (86%). A two-proportion z-test confirmed that this difference is statistically significant ($Z = 3.76$, $p = .000$), suggesting that students were significantly better at identifying the source of content in the logic domain. This discrepancy may be attributed to varying levels of students' familiarity with the subject matter and distinct discursive and lexical features characteristic of each domain. The technical complexity and specialized vocabulary inherent to neurobiology may have hindered the students' ability to detect stylistic cues often associated with AI-generated content, such as overly formal tone or rigid structure.

Descriptive statistics revealed that 62.8% of the students accurately identified the content source providing a baseline for further comparative analysis across academic groups. A statistically significant difference was found between Croatian and Polish students in their ability to recognize AI-generated content. Specifically, 74% of Croatian Computer Science students correctly identified AI-generated content, compared to 57% of Polish students studying Linguistics and Applied Linguistics. This difference was statistically supported by a Mann-Whitney U test ($U = 8224.000$, $p = .004$), suggesting that disciplinary background and academic orientation may significantly influence students' critical awareness of AI-produced material.

The students identified several key features that helped them distinguish AI-generated content from human-written text. AI-generated explanation was typically longer and more detailed, including additional facts, statistics and numerical data that extended beyond a basic definition. For example, AI-generated text frequently provided specific measurements or ranges, such as wavelengths, and included comprehensive, textbook-like descriptions. The language used was generally more formal and academic, employing complex vocabulary and polished sentence structures. Stylistically, the text often contained distinctive markers like the use of em dashes (—), consistent capitalization of technical terms and a tone that sounded highly structured and "encyclopedic," sometimes perceived as less natural or overly professional. In contrast, the content identified as human-created was usually more concise and direct, focusing on the core concept without unnecessary elaboration. Simpler language, occasionally with informal elements or slight imperfections such as lowercase letters where capitals might be expected or shortened scientific terms like "oxy- and deoxyhemoglobin" was used. The human-created text felt more intuitive and accessible, reflecting a conversational or practical style that was easier to understand. The students noted that human explanation sometimes assumed shared knowledge and omitted details considered obvious, giving them a more personal or relatable tone.

C.5 Comparative Evaluation: Neurobiology Task

Unlike with the logic-related excerpt, the Wilcoxon signed-rank test ($Z = -1.476$, $p = .140$) did not reveal a significant difference in the students' perceived difficulty when comparing human- and AI-generated neurobiology-related content. Similarly, no significant differences were recorded in the students' evaluations of usefulness ($Z = -5.230$, $p = .401$), credibility ($Z = .170$, $p = .865$) and engagement ($Z = -1.836$, $p = .066$) between human-created and AI-generated neurobiology-related content. The only recorded difference was observed in examples provided in AI-generated content ($Z = -4.300$, $p = .000$). These results suggest that when it comes to complex, domain-specific topics such as neurobiology, students may find it equally challenging to evaluate and understand content regardless of whether it is human-created or AI-generated. Furthermore, the lack of significant differences in perceived usefulness, credibility and engagement indicates that AI-generated content in this domain may approximate the quality of human-created content from the learners' perspective.

The students were more successful in identifying the human-created explanation in the logic-related excerpt than in the neurobiology-related one. In addition to achieving higher accuracy, they also reported finding the logic-related content easier to comprehend. A Wilcoxon signed-rank test revealed a statistically significant difference in perceived comprehension between the two domains ($Z = -6.773$, $p = .000$), indicating that the students found the logic-related human explanation significantly easier to understand than the neurobiology-related one. They also found the logic-related human explanation more useful ($Z = -7.305$, $p = .000$) and reliable ($Z = -3.099$, $p = .002$). These findings suggest that domain familiarity and content complexity substantially affect students' ability to critically evaluate and comprehend textual information.

Furthermore, the Mann-Whitney U test revealed gender differences in the male students perceiving neurobiology-related human explanation easier to understand ($Z = -2.762$, $p = .006$) and more useful ($Z = -2.310$, $p = .021$) compared to their female counterparts. This suggests that the male students may have greater familiarity or confidence with the content or presentation style of the neurobiology-related material, which could influence their evaluation of its clarity and utility. These findings highlight the potential impact of gender on students' engagement with and perception of academic texts, warranting further investigation into how instructional design can address such differences.

C.6 Content Type Preferences: Neurobiology Task

When the students were asked to indicate their preference regarding the source of the neurobiology-related explanation, 21.5% expressed a preference for human-created content, 31.9% favored AI-generated explanations and the largest proportion, 46.6%, indicated a preference for a combination of both sources. In contrast, preferences for the logic-related explanations differed significantly, namely, 38.3% preferred human-created explanations, 48% opted for a hybrid approach and only 13.8% ($N = 41$) favored exclusively AI-generated content. These results suggest that students perceive the value of AI-generated content differently depending on the academic domain. For neurobiology, a highly specialized and complex subject, nearly one-third of students trusted AI-generated explanations and almost half preferred a combination of human and AI sources, indicating openness to AI assistance but also a desire for human expertise. The relatively low preference for purely human explanations in neurobiology may reflect recognition of AI's ability to handle detailed, data-rich content in technical fields. Conversely, in the logic-related domain (likely more familiar and accessible to students), a higher proportion preferred human-created or hybrid explanations with fewer students fully endorsing AI-generated content. This may reflect greater confidence in human explanations for conceptual and reasoning-based material, where nuance and critical thinking are crucial. Our findings align with Berber Sardinha's (2024) research highlighting the domain-specific nature of human ability to detect AI-generated versus human-authored texts. A multidimensional comparison demonstrated that AI's proficiency in mimicking human writing varies significantly across text types, with AI performing better in creative registers such as poetry but less convincingly in more formulaic genres like article abstracts. This underscores the importance of considering domain and text genre when evaluating AI-generated content detection and suggests that educational interventions should be tailored accordingly to improve critical literacy skills in diverse academic contexts.

C.7 Qualitative Feedback: Human-Created Explanations

The qualitative analysis of the students' responses revealed several prominent themes regarding the perceived qualities of the human explanations provided. The majority of the comments ($N = 71$) emphasized clarity, simplicity and ease of understanding, highlighting that the students valued concise and straightforward language that facilitated comprehension. A substantial subset of the responses ($N = 29$) focused on the professionalism and scientific rigor of the explanations, noting the use of precise terminology, factual accuracy and reliability as important factors influencing their preferences. The human factor and writing style were also frequently mentioned ($N = 18$), with several students appreciating explanations that reflected a human-like tone or a familiar narrative style. Additionally, content-related aspects, including informativeness and the inclusion of practical examples, were noted in 16 responses, underscoring the importance of contextualizing information through real-life applications. Conversely, a smaller portion of students ($N = 13$) reported difficulty in understanding certain explanations, indicating potential challenges related to domain-specific terminology or complexity. Lastly, a group of neutral or non-committal responses ($N = 13$) suggested indifference or uncertainty regarding the explanations provided.

C.8 Qualitative Feedback: AI-Generated Explanations

The analysis of the qualitative feedback on the AI-generated explanation reveals several key thematic categories characterizing the responses. The most prevalent theme, with 76 mentions, highlights the importance of detailed and specific information, where the students valued thorough explanations, inclusion of relevant data and comprehensive descriptions. Closely following, clarity and ease of understanding ($N = 63$) emerged as a critical factor, emphasizing the preference for language and presentation styles that facilitate comprehension, especially for non-experts. The professionalism and technical accuracy of the explanations were noted in 21 instances, reflecting the significance of precision and appropriate scientific terminology for students familiar with the subject matter. Other important categories include the use of examples and real-life applications ($N = 13$), which enhanced understanding by contextualizing concepts in practical scenarios, and engagement and approachable style ($N = 12$), indicating that a relatable and less formal tone contributed positively to reader reception.

Appendix D: AI Usage and Statistical Data

D.1 AI Usage Categorization

Note on terminology consistency: throughout the Results section, “credibility” refers to the perceived accuracy and reliability of a specific text, while “trust” refers to the broader disposition toward AI-generated content as a source type. “Reliability” and “trustworthiness” are used as lay-language equivalents of credibility when describing participants’ evaluations; they are not treated as distinct constructs. Effect sizes for all significant tests are reported at the end of the Results section.

The survey results revealed that 88.3% of the student sample actively integrate AI technologies into their everyday activities. However, the adoption of artificial intelligence in daily life is influenced by a complex interplay of factors, encompassing both motivations for use and reasons for avoidance. The analysis of the students’ responses regarding their use of artificial intelligence tools reveals a nuanced landscape of motivations and concerns, reflecting both practical and ethical considerations. Positive attitudes toward AI use are primarily driven by factors such as efficiency and convenience, with students emphasizing the speed and ease of obtaining information, which aligns with França’s findings (2023) who highlights AI’s role in enhancing research efficiency and data analysis capabilities. Additionally, AI is commonly employed as a study aid, assisting students in understanding explanation or verifying answers. Some students noted that AI has been already embedded in everyday applications, such as search engines or health tracking apps, while others reported using it occasionally or only when necessary, suggesting a selective adoption rather than consistent reliance.

In contrast, 11.7% of the students expressed reasons for not using AI. The most frequently cited barrier was a perceived lack of personal need or relevance ($n = 10$), followed by concerns over misinformation and low trust in AI-generated content ($n = 6$) that were similarly reported by Pitts et al. (2025). Several respondents preferred traditional methods of inquiry, such as conducting their own research or relying on existing knowledge ($n = 8$), underscoring a commitment to independent cognitive engagement. This aligns with Chan and Hu’s (2023) study, which identified a significant cohort of students who view reliance on AI as detrimental to academic skill development. Ethical concerns, particularly relating to environmental impact, copyright infringement and the societal implications of generative AI, were raised by six respondents. Furthermore, philosophical opposition ($n = 4$), privacy concerns ($n = 1$) and accessibility issues ($n = 1$) were also mentioned. A subset of responses reflected a reluctant or selective use of AI ($n = 5$), indicating ambivalence rather than outright rejection.

The categorization of AI usage among the students reveals a multifaceted integration of these technologies into academic and personal contexts. The most prominent uses are content creation for assignments (e.g. essay drafting or presentation design), problem-solving, information analysis and language assistance (e.g. grammar and translation tools). This finding is consistent with the research of Digital Education Council (2024) and Jingwei He et al. (2024). In addition to academic uses, the students also reported employing AI for personal productivity (e.g. task summarization and schedule management) and communication and learning (e.g. educational games and self-improvement tools). While less prominent, AI-assisted shopping, personalized recommendations and smart living applications reflect emerging patterns in AI-assisted consumer behavior. Meanwhile, health, wellness as well as emotional and social support show the students’ growing trust in AI for mental health and companionship needs.

D.2 Extended Cross-National Comparisons

To explore cross-national perceptions, a Mann-Whitney U test was conducted to compare the responses from Polish and Croatian students on perceived difficulty of AI-generated explanation. The results indicated no significant difference between the two groups ($U = 9259.500$, $z = -.973$, $p = .330$) indicating comparable general attitudes toward AI-generated and human-authored content. A comparable pattern emerged in their levels of trust in AI as a source of academic knowledge, with no statistically significant difference between the Polish and Croatian students ($U = 9102.00$, $z = -1.214$, $p = .225$). This similarity in trust levels may reflect a shared digital literacy environment or comparable exposure to AI tools across both educational systems, indicating a consistent cross-cultural perception of AI's role in knowledge delivery.

D.3 Discipline-Based Variation: Extended Analysis

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). 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 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 highlighting how language-specialized students engage more critically with machine-produced texts. These results imply that while most students may perceive AI-generated explanation 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 applied in language-focused academic contexts.

D.4 Effect Sizes for Significant Statistical Tests

Table D.1

Effect Sizes for Statistically Significant Tests

Test	**Comparison**	**Statistic**	***p-value***	**Effect size (*r*)**
Wilcoxon signed-rank	Human vs. AI explanation difficulty (N = 298)	z = -3.263	.001	r = .19 (small)
Mann-Whitney U	Daily users vs. non-users: AI comprehension difficulty	z = -3.492	.000	r =.20 (small)
Mann-Whitney U	Daily users vs. non-users: trust in AI content	z = -6.287	.000	r = .36 (medium)
Mann-Whitney U	Poland vs. Croatia: AI difficulty (no sig.)	z = -.973	.330	n/a
Mann-Whitney U	Poland vs. Croatia: AI trust (no sig.)	z = -1.214	.225	n/a
Kruskal-Wallis	Discipline-based variation, Polish sample	$X^2(6)$ = 12.152	.002	η^2 = .06 (small-medium)