

HUOM! Tämä on alkuperäisen artikkelin rinnakkaistallenne. Rinnakkaistallenne saattaa erota alkuperäisestä sivutuksestaan ja painoasultaan.

PLEASE NOTE! This is a self-archived version of the article that may differ from the final publication by pagination and typography.

Viittaa alkuperäiseen lähteeseen:

Cite the final publication:

Alamäki, A., Khan, U. A., Kauttonen, J., & Schlögl, S. (2024). An Experiment of AI-Based Assessment: Perspectives of Learning Preferences, Benefits, Intention, Technology Affinity, and Trust. *Education Sciences*, 14(12), 1386. <https://doi.org/10.3390/educsci14121386>

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).



Article

An Experiment of AI-Based Assessment: Perspectives of Learning Preferences, Benefits, Intention, Technology Affinity, and Trust

Ari Alamäki, Umair Ali Khan, Janne Kauttonen and Stephan Schlögl

Special Issue

How Artificial Intelligence Can Enhance Education: Current Practices and Challenges


Edited by

Prof. Dr. Julie Delello



Article

An Experiment of AI-Based Assessment: Perspectives of Learning Preferences, Benefits, Intention, Technology Affinity, and Trust

Ari Alamäki ¹, Umair Ali Khan ^{1,*}, Janne Kauttonen ¹ and Stephan Schlögl ² 

¹ RDI & Competences, Haaga-Helia University of Applied Sciences, Ratapihantie 13, 00520 Helsinki, Finland; ari.alamaki@haaga-helia.fi (A.A.); janne.kauttonen@haaga-helia.fi (J.K.)

² Department of Management, Communication & IT, MCI Management Center Innsbruck, 6020 Innsbruck, Austria; stephan.schloegl@mci.edu

* Correspondence: umairali.khan@haaga-helia.fi

Abstract: The rising integration of AI-driven assessment in education holds promise, yet it is crucial to evaluate the correlation between trust in general AI tools, AI-based scoring systems, and future behavioral intention toward using these technologies. This study explores students' perspectives on AI-assisted assessment in higher education. We constructed a comprehensive questionnaire supported by relevant studies. Several hypotheses grounded in the literature review were formulated. In an experimental setup, the students were tasked to read a designated chapter of a paper, answer an essay question about this chapter, and then have their answers evaluated by an AI-based essay grading tool. A comprehensive data analysis using Bayesian regression was carried out to test several hypotheses. The study finds that remote learners are more inclined to use AI-based educational tools. The students who believe that AI-based essay grading is less effective than teacher feedback have less trust in AI-based essay grading, whereas those who find it more effective perceive more benefit from it. In addition, students' affinity for technology does not significantly impact trust or perceived benefits in AI-based essay grading.



Citation: Alamäki, A.; Khan, U.A.; Kauttonen, J.; Schlögl, S. An Experiment of AI-Based Assessment: Perspectives of Learning Preferences, Benefits, Intention, Technology Affinity, and Trust. *Educ. Sci.* **2024**, *14*, 1386. <https://doi.org/10.3390/educsci14121386>

Academic Editors: Julie Delello and Rochell McWhorter

Received: 29 September 2024

Revised: 9 December 2024

Accepted: 12 December 2024

Published: 17 December 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: artificial intelligence; AI-based assessment; technology affinity; trust; higher education; EdTech

1. Introduction

Due to its great potential to transform educational practices, artificial intelligence (AI) is currently being researched from several perspectives in education [1,2]. However, there is still limited evidence about the practical results of AI adoption in education [3] and the perceptions and expectations of first-year students related to AI tools [4]. Focusing on first-year students is particularly crucial because their experiences can provide valuable insights into how novices interact with and perceive AI tools within an educational context. Although automated essay grading systems have been in the interest of researchers for a long time [5,6], there are still few studies on students' experiences where students have interacted with AI-based essay grading systems in authentic learning and teaching situations. AI-based knowledge assessment is a crucial part of AI in education [7] since formative assessment and informal assessment of students' own competencies are important activities in education [8,9]. However, Gao and colleagues [5] assert that a relatively small number of automated assessment systems currently support the enhancement of complex thinking and reasoning through feedback.

The utilization of AI for automating learning task assessments, including scoring and provision of qualitative feedback, is one of the key applications of this technology in the education sector. Such automation is intended to save time, facilitate teachers' routine tasks, and increase assessment consistency [6]. At the same time, it provides students with

instant feedback on their assignments and bypasses the often-long waiting times associated with teacher-provided feedback.

Despite these advancements, our understanding regarding student sentiments toward such instructional settings, where AI has been employed for grading and assessing their assignments through written feedback, remains limited. Hence, it becomes important to (i) explore methods of building trust in educational AI solutions; (ii) enhance our understanding of students' intentions concerning the use of AI grading systems; and (iii) examine how various background variables among students may influence their experiences and assumptions related to AI.

We are aware that attitude and trust towards digital learning environments, coupled with perceived usefulness, can stimulate improved learning behavior among students [10]. However, there is a clear need for more research into student experiences when interacting with AI-based learning solutions.

To this end, our study primarily investigates the factors influencing the experiences and perceived behavioral intentions related to the use of an AI-based essay grading system in a real-world educational setting. To operationalize this investigation, students were tasked with writing a short essay. Their written work was then evaluated by an AI-driven essay grading tool. Following this process, students completed a survey designed to investigate their personal experiences, level of trust in the tool, perceptions of its advantages, and their intention to continue using such an AI feedback mechanism.

The primary question guiding our study was:

How do students' characteristics, preferences, and technology affinity affect trust, perceived benefits, and behavioral intention regarding the adoption of AI and AI-based essay scoring systems?

Understanding the influence of student demographics on trust in AI-based essay scoring systems, the perceived benefits of AI-based assessments, and the intention to use AI educational tools is important for two reasons. First, demographic factors such as age, gender, educational background, and technological proficiency can significantly affect how students perceive and interact with AI technologies in educational settings. For instance, a study [11] highlights that teachers' adoption of AI in classrooms is influenced by age, gender, teaching experience, and academic discipline, suggesting that similar demographic variables may impact students' engagement with AI tools.

Second, research indicates that individuals with higher self-efficacy and understanding of AI are more likely to trust and perceive benefits from AI educational technologies [12]. This highlights the importance of considering demographic variables, as they often correlate with varying levels of technological proficiency and familiarity with AI.

Given the limited research specifically addressing the intersection of student demographics and their perceptions of AI-based educational assessments, our study aims to fill this gap. Therefore, we seek to provide insights that can inform the development of AI tools and educational practices that are responsive to the diverse needs and backgrounds of students.

Our report on this study is organized as follows: Section 2 focuses on a review of related work, while Section 3 elaborates on the theoretical framework for the study and our hypotheses. Subsequently, the methods, procedure of the experiment, and data analyses are described in Section 4. Section 5 then reports on the results obtained from our questionnaire, and Section 6 provides a comprehensive discussion of their relevance. Finally, Section 7 summarizes the study's contributions, underscores its theoretical and managerial implications, and outlines the limitations of our work.

2. Related Work

2.1. AI-Based and Automated Essay Scoring Systems

An automated essay-scoring application is a computer-based assessment system that automatically evaluates students' written responses [6]. In this, AI differs from conventional computer programs because AI technologies are based on autonomous decisions,

whereas conventional computer programs are controlled by humans [13]. Aligned with these definitions, an AI-based essay scoring system also automatically evaluates written responses, yet it uses autonomous analytics, such as machine learning. Humans need to configure or hard-code conventional computer-based assessment systems to run automated scoring mechanisms such as assessing quizzes or multi-choice learning tasks. AI-based essay grading systems are trained with a model response to detect, automatically recognize, and correct pieces of knowledge from students' written essays. That is, AI-based essay scoring systems evaluate the relevance of content in various computing methods, such as statistical methods, classification methods, or neural network approaches [6]. In addition to scoring the essay response, providing automatic feedback for students on how to improve the essay is also an important feature of a modern essay scoring system [14]. Continuous feedback about student's performance relates to the formative assessment that aims to support students' learning processes [15,16]. Thus, AI-based essay scoring as a part of formative assessment refers to situations where AI-based assessment supports students' learning.

In developing AI-enhanced learning practices and instructional design in higher education, we need a deeper understanding of students' expectations and experiences about the use of educational AI. Since AI-based essay scoring systems automatically evaluate and score students' answers, students' trust in these systems is critical. Traditionally, students have mainly trusted a teacher's assessment, although there is a variation between disciplines and school subjects. For example, the knowledge content of mathematics and science is more objective and widely accepted than applied knowledge, such as the knowledge related to business development or marketing studies. This causes additional challenges for AI-based essay scoring systems. Assessing the relevance of content in the students' essays is a challenging task for computer systems, even when utilizing AI-based solutions [6].

2.2. Individual Characteristics and Technology Affinity of Students

Individual characteristics influence students' behavior [17], and students exhibit various behavioral patterns in AI-enhanced learning environments [18]. Thus, the attributes of an AI-enhanced learning environment are insufficient to explain the effectiveness of learning and teaching. Prior research points out that technological application alone does not cause learning but is rather an enabler for content sharing or interaction that makes students think [19]. Also, a certain AI anxiety may cause negative affection in students [13]. Consequently, students' characteristics must be understood in order to explain the effects of AI-based essay grading systems. That is, students can have different relationships with the same AI-enhanced solution [20]. In other words, different cognitive and emotional processes drive students' experiences while using AI-based essay scoring systems by creating different relationships with them. To this end, learning psychology literature [21,22] emphasizes the constructive concept of learning, which means that students process new information based on their prior understanding and experiences. In addition to individual value preferences and affective reactions, many other factors, such as demographics and study preferences, contribute to user experiences and preferences while interacting with AI-enhanced learning and teaching. In order to frame hypotheses regarding individual characteristics impacting AI use, we considered demographic influences, students' study preferences, and preferences concerning teacher- or AI-assisted teaching. Given the absence of previous studies investigating such correlations, our hypotheses about this field have both a confirmatory and exploratory nature. The following sections explain the process of formulating these hypotheses in more detail.

3. Theoretical Framework and Hypotheses

In order to formulate the hypotheses for this study, we selected four predictor groups, i.e., general trust in AI applications, technology interaction, study preferences, and demographic variables. The target groups included those who preferred AI over teachers,

those who intended to use AI tools, those who perceived benefits, and specific trust. In the following sections, we discuss the formulation of the hypotheses we tested in this study.

3.1. Demographic Influence

The relationship between students' demographics and their trust in, and perceived benefits of, AI-based essay grading tools remains underexplored. However, existing research suggests that demographic factors such as age, gender, and educational background may influence trust, behavioral intentions, and perceived benefits regarding AI technologies in general. To better understand these dynamics in the context of AI-based essay grading tools, we examine the following associations.

3.1.1. Age and Trust in AI-Based Essay Grading

Demographics are an important factor in attitudes toward AI, although there are conflicting findings regarding the relationship of attitudes towards AI, age, and gender [23]. Younger individuals, often referred to as digital natives, have grown up in an environment saturated with technology. According to a study by Rogers [24], younger people are typically early adopters of new technologies due to their openness to change and higher exposure to technological advancements. This familiarity can lead to increased trust in AI tools, and they may also exhibit greater trust in AI-based essay grading systems. For instance, studies have shown that consumers under the age of 45 in the United States have a relatively higher trust in AI [4,25]. A similar trend was found in Australia, where 42% of Gen X and Millennials were found to have higher trust in AI compared to 25% of the older generation [26]. The same study [26] further found that this trend is global, where younger people with university education are more accepting of AI [26]. These studies suggest that age-related differences in technology exposure and adoption could play a significant role in shaping trust in AI. Hence, a hypothesis may be developed to investigate the association between age and trust in AI-based essay grading.

3.1.2. Gender and Trust in AI-Based Essay Grading

The relationship between gender and trust in AI varies across studies. A study explores the gender differences in technology adoption and trust through the lens of the gender schema theory [27], which posits that societal norms and stereotypes influence individual behavior and attitudes. Some studies suggest that males tend to show more trust in AI than other genders [28,29], potentially due to higher engagement with technology and STEM fields [30,31]. However, the global study on trust in AI [26] indicates that there are no meaningful differences across men, women, and other genders in terms of trust, acceptance, or emotions toward AI. Nevertheless, in a few countries such as the USA, Singapore, Israel, and South Korea, men showed a slightly higher level of trust or acceptance of AI and reported more positive emotions towards AI compared to other genders. In a similar way, Omrani and colleagues [32] found that men trusted more in AI than women. Given these mixed findings, it is pertinent to investigate if an association between gender and trust in AI-based essay grading exists [26].

3.1.3. Demographics and Perceived Benefits of AI-Based Essay Grading

Educational background, particularly in technical fields, impacts AI literacy and, consequently, trust in AI tools. Students with technical education may perceive greater advantages due to their familiarity and comfort with technology. Conversely, students from non-technical backgrounds might be more skeptical, perceiving fewer benefits. Research by Castillo-Acobo and colleagues [33] indicates that age and field of study influenced AI usage in classrooms, with younger students and those in STEM fields being more likely to use AI. The study by Davis [34] emphasizes that perceived benefits and ease of use influence an individual's acceptance of technology. Therefore, it could be interesting to explore whether age and educational programs in combination also share students' perceived benefits of AI-based essay grading.

3.1.4. Demographics and Behavioral Intention to Use AI-Based Educational Tools

The behavioral intention to learn AI has also been explored in various educational contexts. Studies show that students' intentions are often shaped by factors such as AI literacy, self-efficacy, and social influences, which can be impacted by demographics like age or educational context [10]. Hence, it is worth exploring whether demographic factors (including age and educational programs) also shape students' behavioral intention to use AI-based educational tools in the future.

3.1.5. Educational Programme and Trust in the AI-Based Essay Grading Tool

Students with a technical study background or prior experience in AI have been demonstrated to show higher AI literacy [35]. Hence, they have a better understanding of the nuances of AI tools and their decision-making because of the higher exposure to AI and machine learning-related courses. A study shows that students with a higher self-reported understanding of AI hold more positive thoughts about integrating AI into their classrooms [36]. Similarly, another study by Long and Magerko [37] found that individuals with higher AI literacy demonstrated more positive attitudes toward AI integration in educational settings. Additionally, the study by Venkatesh et al. [38] suggests that facilitating conditions, such as prior knowledge and experience, enhance the likelihood of technology adoption. Applying these findings to AI-based essay grading tools, it is plausible that students enrolled in information technology or digital education programs may exhibit higher trust in such tools. Therefore, investigating the association between a student's educational program and their trust in AI-based essay grading systems is warranted.

Consequently, we may formulate the following hypotheses regarding demographic influence on technology affinity and trust in AI:

- H1.** *Younger students (under 40) have higher trust in AI-based essay grading than older students.*
- H2.** *Male students tend to show more trust in AI-based essay grading compared to other genders.*
- H3.** *There is a significant association between students' demographics and the perceived benefits of AI-based essay grading.*
- H4.** *There is a significant association between students' demographics and behavioral intention to use AI-based educational tools in the future.*
- H5.** *Students of IT or other digital education programs have higher trust in the AI-based essay grading tool than the students of other degree programs.*

3.2. Study Preferences

While no studies directly confirm any associations between study habits and trust in AI-based grading, the studies on AI-based tools in online learning (remote education) show an increasing trend [39], indicating that remote learners are more inclined to use such technologies. AI-based tools can be used to provide personalized learning experiences for students based on their individual needs, strengths, and weaknesses [40], which are particularly important for students studying alone. Recent studies have also shown promise in supporting learners' self-regulation in online learning by measuring and augmenting self-regulated learning [41]. Zimmerman [42] suggests that self-regulated learners actively control their learning processes through goal setting, self-monitoring, and self-assessment. AI-based essay grading tools align well with self-regulated learning by offering instant feedback, allowing students to iteratively improve their work autonomously. Another work by Wei and colleagues [43] shows that personalized AI-driven tools are effective for providing appropriate learning resources in remote education and enhancing students' learning outcomes.

This made us formulate the following hypotheses:

H6. *Students who prefer to study remotely have a higher behavioral intention to use AI-based educational tools.*

H7. *Students who prefer to study alone perceive more benefits from AI-based essay grading.*

H8. *Students who prefer to study alone have more trust in AI-based essay grading.*

3.3. Preferences Regarding Teacher-Led vs. AI-Led Instruction

An AI-based assessment offers several advantages to students, including instant feedback and unbiased grading. It also aims to decrease the workload and fatigue of teachers, enabling them to concentrate on more advanced pedagogical activities instead of routine and procedural tasks [44]. The study presented by Escalante et al. [45] shows that half of the students preferred AI-generated feedback, and they mentioned clarity and specificity of the feedback as benefits compared to the feedback given by a teacher. Research has furthermore shown that AI-based essay scoring systems not only enhance the efficiency of essay assessment but also contribute to more objective scoring methods [46]. However, the literature does not offer detailed insight into students' perceptions of AI-driven scoring systems. A study by Stoica [47] concludes that AI-based essay scoring is not valued more than that of humans. This is confirmed by another study, which suggests that even when explainable AI is used in essay grading, it is not better than human grading for students. Another study indicates that even with the transparency offered by explainable AI, students do not regard it as surpassing human assessment in essay grading [47]. These results suggest that the main cause of this attitude was a lack of trust in the AI-based essay grading system. Hence, one may infer that this behavior might be consistent with less perceived benefits from AI-based grading as well as lower behavioral intention to use AI-based educational tools. A study by Nazaretsky and colleagues [48] examined teachers' trust in AI-based educational technology and identified key factors such as perceived benefits, lack of human characteristics, and transparency as essential for building trust. This suggests that teachers' assessments may still be favored over AI-based assessments. Further experiments with K-12 science teachers demonstrated that providing explanations of AI decision-making processes and comparing AI with human performance can boost trust and encourage the use of AI-powered tools for assessments [49], similar to how teacher explanations enhance assessments. Additionally, researchers [50] highlighted that teachers' prior knowledge and beliefs significantly influence their trust in AI-based assessments, which can, in turn, shape students' perceptions of such technologies.

This leads to the formulation of the following hypotheses:

H9. *Students who believe that AI-based assessment is more effective than teacher feedback have more trust in AI-based essay grading.*

H10. *Students who believe that AI-based assessment is more effective than teacher feedback perceive more benefits from AI-based essay grading.*

H11. *Students who believe that AI-based assessment is less effective than teacher feedback have a lower behavioral intention to use AI-based educational tools.*

3.4. Affinity for Technology Interaction (ATI)

Affinity for technology interaction (ATI) refers to a person's tendency to actively engage in intensive technology interaction or to avoid it [51]. To the best of our knowledge, there are no studies in the field of education that directly investigate the association of ATI and AI trust and perceived benefits. Relevant studies in other domains regarding ATI and AI trust and perceived benefits also report mixed results. A recent study [52] reveals a positive correlation between ATI and trust in AI through a survey that asked participants how they perceive 38 statements about AI in different contexts (personal,

economic, industrial, social, cultural, and health). However, some studies that focus on specific AI applications, such as legal decision-making [53] and intelligent virtual assistants [54], find no significant correlation between ATI and trust in AI. On the other hand, a study by Buck et al. [55] suggests that the interest in using AI-based applications also depends on the general affinity for technology.

Research in the education field indicates that teachers with greater self-efficacy and understanding of AI-based educational technology tend to perceive more benefits, have fewer concerns, and exhibit higher levels of trust in these tools [12]. This attitude can also influence students, as teachers often serve as role models in adopting and utilizing new technologies. In another study, Delcker et al. [4] found that students' intention to use AI tools is influenced by their perceived skills, knowledge, and attitudes. Additionally, students' curiosity about new technology encourages hands-on testing, which can deepen their practical understanding of AI tools and their potential benefits. However, none of these studies specifically addresses ATI in its true sense. Hence, it is worth exploring whether students' ATI has any correlation with their trust in AI-based essay grading, perceived benefits of AI-based essay grading, and behavioral intention to use AI-based educational tools.

This leads to the following exploratory hypotheses:

H12. *Student's affinity for technology interaction does not significantly influence their trust in AI-based essay grading.*

H13. *Student's affinity for technology interaction does not significantly influence their perceived benefits of AI-based essay grading.*

H14. *Students with a higher affinity for technology interaction have higher behavioral intention to use AI-based educational tools.*

3.5. The Relationship of Trust, Perceived Benefits, and Intention to Use AI

AI-enhanced educational solutions can enrich learning and teaching [56]. Simultaneously, the adoption of AI-enhanced educational solutions in learning and teaching may also influence negative experiences and outcomes, such as privacy concerns and ethical concerns [13,57,58]. Trust is also an essential factor for AI adoption in education [59]. The study by [60] defines trust as the extent to which individuals believe technology is credible, reliable, and secure. The existing studies show that students may also feel anxious and less confident when learning with AI [1]. It is important to study trust in developing AI-enhanced learning and teaching, as the simultaneous interaction of physical, virtual, and social worlds in learning and teaching makes the learning environment a complex phenomenon. Socio-constructive knowledge creation in connection with practical hands-on activities [61], the ability to mix virtual and physical environments [62,63], and competencies to integrate digital tools into teaching arrangements [64] cause several interrelated variables that affect the students' cognitive preferences, thinking processes and affective experiences. This way, they also affect perceived trust in AI-based essay scoring systems.

We already know that affective information creates more effective results than rational, cognitive information alone [65,66]. For example, positive media content that evokes high-arousal emotions goes more viral than negative content [67]. Emotion is not a concrete object or media element in a digital environment that just flows out to the minds of individuals. Human behavior is complicated as humans' cognitive experiences play a critical role in stimulating affective reactions [68]. Thus, cognitive and affective experiences cannot be separated from each other. Attitudes of individuals influence intention, which affects human behavior [69]. Aligned with this, students' knowledge about AI and general AI anxiety affects their behavioral intention toward learning AI [10].

The interaction with digital solutions stimulates the direction of affective experiences [70]. Similarly, students process new information using an AI-based essay scoring

system, but they also often unintentionally form an affective state and willingness regarding the trust, perceived benefits, and intention to use it in the future. Students' experiences related to digital tools are important because they support a learner's active participation, inner motivation, and first-order cognitive and affective experiences [71,72]. These factors, in connection with an AI-enhanced learning environment, optimally support students' learning according to the constructivist learning approach. The cognitive appraisal theory [68] points out that students cognitively observe and interpret digital feedback and assessment and reflect on their personal experiences, which affect their affective feelings. Hence, positive affective experiences influence students' behavioral intention to use similar AI-enhanced solutions. General trust in AI applications is very important for their widespread adoption and acceptance [26]. Moreover, the perceived benefits of AI applications are also related to the trust in AI applications [26]. This is further supported by findings from Hall et al. [73], who identify trust as one of the main key factors for the acceptance of AI-based essay grading tools. Consequently, we may hypothesize:

H15. *Students who have higher trust in AI applications also have higher trust in AI-based essay grading.*

H16. *Students who have higher trust in AI applications have higher perceived benefits of AI-based essay grading.*

H17. *Students who have higher trust in AI applications have higher behavioral intention to use AI-based education tools.*

4. Methods and Materials

4.1. Participants

The data were collected from two different Universities of Applied Sciences situated in Austria and Finland, comprising a total of 158 responses. Out of these responses, 116 were fully completed, while the remaining 42 were partially filled out. For those 42 subjects, we could use 38 in factor analysis and 26 in regression analysis (see details in Results). The majority of participants identified as female, accounting for 55.1% (88 out of 158), and the average age category was predominantly within the range of 21 to 30 years old (with a standard deviation of 0.62). Our participant pool consisted not only of students pursuing their first undergraduate degree but also career changers and continuous learners from various age groups. At the time of our study, a significant portion, i.e., 71.5% (113 out of 158), were enrolled in Business programs.

Interestingly, around 65.2% (103 participants) expressed their preference for studying alone rather than engaging in small group studies, indicating a tendency towards individual learning practice among our sample group. Despite the increasing popularity of remote or distance learning, more than half, i.e., 55.1% (87 participants), expressed their preference for on-campus study, highlighting a majority preference for traditional learning settings within our participant group.

4.2. AI-Based Essay Scoring Setup

For this experiment, we utilized the proprietary software "Eximiatutor 1.0", developed by Eximia AI Technologies Ltd. in Espoo, Finland. This AI-driven software specializes in scoring and assessing essays. The experiment was conducted using a separate instance of this system, operating as a cloud service that could be controlled and accessed via a web browser. This ensured that the system was free from any pre-existing data or information related to the study's respondents, thereby preventing potential bias in the results. Eximiatutor operates as an automated text-based assessment platform that produces both written and numeric feedback from students' text responses. The software presents a list of correct responses in textual form, alongside their corresponding percentage accuracy when compared to the model answer on which it has been trained.

For this experiment, we asked the respondents to read Section 5 of the article “Business Impacts of Technology Disruption—A Design Science Approach to Cognitive Systems’ Adoption within Collaborative Networks” written by [74] and subsequently, respond to an essay-type question related to the article’s content. The selected article was deemed appropriate as it mirrors topics currently taught to undergraduate students in business administration. This approach allowed students to first experience how the tool evaluated their work and then ground their perceptions and trust in actual engagement rather than theoretical assumptions. By engaging with the tool in a real-world context, students gained practical insight into its capabilities and limitations. This practical engagement allowed them to reflect and provide informed feedback on their trust, perceived benefits, and future use intentions.

To grade the essay responses accurately, our AI-based grading system was trained on the same article. A human who trained the AI model read the article and selected the sentences that needed to be included in the perfect essay-based answer. In addition to the selected sentences from the article, he also created similar sentences to train the AI model to understand neighboring concepts within the same topic. Finally, the AI system achieved 100% coverage of the required factual issues. The training dataset was based on the facts in the article, but potential biases and limitations were related to the neighboring concepts or synonyms that the AI model could not recognize, such as badly misspelled words and uncommon neighboring terms. Evidently, these did not exclude the entire correct sentence from the space of proper concepts and their relationships but probably weakened its grading. Before initiating the actual experiment, we assessed the accuracy and relevance of the feedback produced by the AI system, along with its scoring results. The AI system employs machine learning-based knowledge graph technologies as part of its natural language processing (NLP) capabilities. During the training of the AI system’s model, a human-verified the results to minimize biases.

4.3. Procedure of Research Experiment

We engaged volunteer lecturers who were keen to incorporate this real-life learning task and research experiment into their undergraduate courses. As such, the experiment was framed as an actual learning assignment for students, typically taking between 30 and 60 min to complete. Importantly, it was emphasized that the experiment would not have any bearing on students’ grades.

Before the students could participate in the research experiment and use the AI-based essay grading system, they were presented with detailed information about the research. Following this, they were asked to give informed consent to participate by choosing “Yes”. Upon agreement, they gained access to the main page of the AI-based essay scoring system. The following instructions for students were devised and displayed on the main page of the system:

1. Please read Section 5 of this article (note: please open it in a separate browser tab) <https://hal.inria.fr/hal-02191186/document> (accessed on 11 December 2024).
2. Please answer the following question: “In which different ways may big data and artificial intelligence (e.g., cognitive systems) change businesses and industries?”
3. Please answer this question by writing a short essay (50–100 words). Start your answering process by clicking on “Open answer form”. On submission of your answer, you will receive feedback and a respective assessment score from an AI grading tool. Note: The feedback and score have NO IMPACT on any of your grades or assessments in class! [Open answer form]
4. After having received your feedback and score from the AI tool, please complete the following survey: [Link to questionnaire]

Upon completion of their essay responses, students submitted their work to the system, which in turn provided textual feedback and a numerical assessment of their performance. The feedback included a student’s original answer and the identified areas or points that may have been overlooked. The system also generated an assessment score as a

percentage, determined by the ratio of accurate to missed information. In addition, students were given access to view the example answer. Thus, the AI system presented students with an example answer after they had submitted their essays and received a numerical assessment along with a list of missing facts. The example answer, including all required facts, was as follows: “Technology disruption relates to three significant concepts: big data, artificial intelligence and cognitive systems. These disruptions (digital development solutions) form a developmental continuum. Technology disruptions start to develop from companies’ business processes and eventually affect and subsequently change an industry’s ecosystem. Business development as technology development is affected by four factors: (1) the development of the business processes takes place supported by the use of big data; (2) the position in the value chain changes according to a higher level of development, such as artificial intelligence and big data; (3) companies notice the impact of how digital development changes the business environment; (4) cognitive systems change the entire business ecosystem. Typically, a company with a strong vision sets out to develop while others follow, so that eventually the entire ecosystem changes.”.

They were then prompted to proceed to the study’s questionnaire via a clickable link.

4.4. Design of the Questionnaire

The questionnaire contained 8 groups (see Table 1). The demographic group comprised ‘gender’, ‘age’, and ‘educational programme’. ‘Study preferences’ explored students’ preferred study mode and location. The ‘AI-based tool vs. teacher feedback’ group assessed opinions on the effectiveness of AI for essay feedback. The ‘Affinity for technology interaction’ measured engagement with and interest in technology. The ‘trust in AI applications’ looked at the ease of trust and feelings towards AI. The ‘trust in the AI-based essay assessment tool’ evaluated its reliability and the trustworthiness of its feedback. The ‘perceived benefits’ group considered how the AI tool might enhance learning and provide feedback. The ‘behavioral intention to use AI-based educational tools’ reflected the likelihood of future use and recommendation of the AI tool.

The groups were adopted from different studies. Question 3 on AI vs. teacher’s feedback was adopted from ‘satisfaction of use’ for comparing traditional teaching methods [51,75]. The group of ‘affinity for technology interaction’, comprising 9 questions based on a 6-point Likert scale, was adopted from the study by Franke et al. [51], where the tendency to actively engage in intensive technology interaction was analyzed. The group ‘trust in AI applications’ and ‘trust in AI-based essay tool’, comprising 3 and 4 questions, respectively, were adopted from a study by Cheung and To [76], which previously found an influence of the propensity to trust on mobile users’ attitudes toward in-app advertisements. The group of ‘perceived benefits’, comprising 5 questions, was adopted from a study on the technology acceptance model in examining students’ behavioral intention to use an e-portfolio system [77]. The behavioral intention group was adopted from a study that studied the correlation between behavioral intention and future transactions in a marketplace with several other factors [78].

Table 1. Structure of the questionnaire.

No.	Group	Data Type	Scale	Questions	Reference
1.	Demographic	Categorical	-	Age: 1 = Under 20, 2 = 21–30, 3 = 31–40, 4 = over 40 Gender: 1 = Female, 2 = Male, 3 = Other or skipped Educational Programme: 1 = IT or Digital, 2 = Business, 3 = Hospitality or Tourism, 4 = Others	Self-developed
2.	Study preferences	Categorical	-	1. How do you prefer to study? 2. Where do you prefer to study?	Self-developed

Table 1. Cont.

No.	Group	Data Type	Scale	Questions	Reference
3.	AI vs. teacher feedback	Ordinal	Scale 2	I think that this AI-based essay tool is more effective than a teacher's feedback	Adapted from (Sung and Mayer, 2012) [75]
4.	Affinity for technology interaction (ATI)	Ordinal	Scale 2	<ol style="list-style-type: none"> 1. I like to occupy myself in greater detail with technical systems. 2. I like testing the functions of new technical systems. 3. I predominantly deal with technical systems because I have to. 4. When I have a new technical system in front of me, I try it out intensively. 5. I enjoy spending time becoming acquainted with a new technical system. 6. It is enough for me that a technical system works; I don't care how or why. 7. I try to understand how a technical system exactly works. 8. It is enough for me to know the basic functions of a technical system. 9. I try to make full use of the capabilities of a technical system. 	Adapted from (Franke et al., 2019) [51]
5.	Trust in AI applications	Ordinal	Scale 1	<ol style="list-style-type: none"> 1. It is easy for me to trust AI applications 2. My tendency to trust AI applications is high 3. I tend to trust AI applications even though I have little knowledge about how they work 	Adapted from (Cheung and To, 2017) [76]
6.	Trust in AI-based essay tool	Ordinal	Scale 1	<ol style="list-style-type: none"> 1. The results of this AI-based essay tool are believable 2. The results of this AI-based essay tool are accurate 3. I trust the results of this AI-based essay-tool 4. The results of this AI-based essay tool are reliable 	Adapted from (Cheung and To, 2017) [76]
7.	Perceived benefits	Ordinal	Scale 1	<ol style="list-style-type: none"> 1. Using the AI-based essay tool could enhance my effectiveness in learning 2. Using the AI-based essay tool could improve my studies 3. Using the AI-based essay tool could increase my productivity in my studies 4. Using the AI-based essay tool could help me to accomplish studies more quickly 5. I found using the AI-based essay tool useful 	Adapted from (Shroff et al., 2011) [77]
8.	Behavioral intention to use AI-based educational tools	Ordinal	Scale 1	<ol style="list-style-type: none"> 1. Given the chance, I predict I would be happy to use AI-based educational tools in the future 2. It is likely that I will have the possibility to use AI-based educational tools in the near future 3. If I'm given the opportunity to use AI-based educational tools, I intend to do so in the future 	Adapted from (Pavlou and Gefen, 2004) [78]

Scale 1: 1 = strongly disagree, 2 = disagree, 3 = neither, 4 = agree, 5 = strongly agree (Likert scale). Scale 2: 1 = completely disagree, 2 = large disagree, 3 = slightly disagree, 4 = slightly agree, 5 = large agree, 6 = completely agree.

4.5. Methods

To create factor constructs, we applied Lavaan (v0.6-17) [79], a software package for R, for our confirmatory factor analysis (CFA). We applied the “cfa” function with the Maximum Likelihood (ML) estimation algorithm to estimate the parameters of our model with missing values. The reliability of the model was evaluated by calculating the Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), and root-mean-squared error. Internal consistency of factors was evaluated by computing Omega, Alpha, and average explained variance. Additionally, we also computed standardized factor loadings to find the correlation between the observed variables and latent variables. Different factor constructs were evaluated to find the one that fit the data best. MacCallum et al. (1999) [80] demonstrated that samples as small as 100 can be sufficient for CFA if the model is well-specified, factors are well-defined with multiple high-loading indicators, and communalities (i.e., variance explained) are high. Similarly, ref. [81] noted that smaller samples might be adequate when the CFA model is not overly complex, and the data exhibit strong factor loadings. Ref. [82] recommend using a variables-to-factors ratio of at least 7 for factor analysis.

After fitting factors, the investigation proceeded with regression analysis, which involved four distinct models corresponding to different dependent outcomes: (1) preference for AI over teacher feedback, (2) behavioral intention to use AI-based educational tools, (3) perceived benefits of AI-based essay grading, and (4) trust in the AI-based essay grading tool. These dependent variables, shown in Figure 1, represent key aspects of students’ engagement and perception of AI in educational contexts. Independent variables included in these models encompassed general trust in AI applications, an affinity for technology interaction (ATI), study preferences (preferred learning style and location), and demographic variables (age, gender, educational program). Variables associated specifically with the AI-based essay-grading tool are highlighted in red in Figure 1. These independent variables were analyzed for their impact on each dependent outcome using Bayesian regression techniques, capturing both linear and interaction effects where applicable.

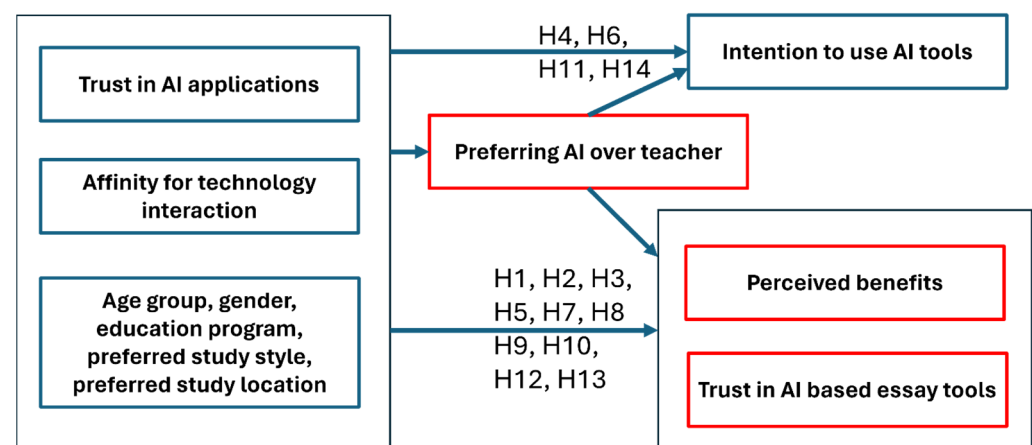


Figure 1. Diagram of regression models and related hypotheses. Four separate regression models were fitted with outcomes preferring AI over teacher, intention to use AI tools, perceived benefits, and trust in AI-based essay tool. Variables marked with red were associated with the essay-grading tool.

In Bayesian regression, the posterior distributions of model parameters were estimated using the Markov chain Monte Carlo method with the brms R package [<https://github.com/paul-buerkner/brms>] (accessed on 11 December 2024) version 2.20.3 [83]. We chose Bayesian regression over classical (frequentist) regression because it allowed us to incorporate prior information to enhance parameter estimation with our limited sample size, provided greater flexibility in specifying and comparing suitable probability distributions and predictor sets, and facilitated the fitting of more robust models that better capture the underlying data structure. The final models were searched among 40 models for factor outputs, each with both normal and skewed normal dis-

tributions (i.e., a total of 80 models per output), while 41 cumulative ordinal models (logit link) were fitted for “Preferring AI over teacher” ordinal output. Variables of age group and preferring AI over teacher were modeled as monotonous variables when used as predictors in models [83]. We used weakly informative zero-mean normal priors for regression coefficients (SD 10); otherwise, priors were kept as built-in defaults of brms. The weak prior was chosen according to recommendations to improve convergence and model identifiability, as well as to reduce the risk of overfitting (see <https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>). Sensitivity analyses were performed to verify that our results remained consistent across different prior configurations (SDs between 5 and 20).

For each model, we constructed five chains of 9000 steps, including 4000-step warm-up periods; thus, a total of 20,000 steps were retained to estimate posterior distributions for each of the five responses. Convergence of the chains was verified by visual inspection of prediction distributions (posterior predictive check) by comparing the observed data to simulated data from the posterior predictive distribution. The checks demonstrated that our chosen models adequately captured the underlying data structure. We also ensured that the potential scale reduction factor R on split chains was equal to 1.0, with estimated effective bulk and tail samples of at least 10,000 for each parameter. Among all model candidates, we applied an approximate leave-one-out cross-validation (LOO) scheme [84,85] that ranks models using their fitted posterior probabilities. The model with the highest probability (within 1 standard error) and complexity was chosen for each of the four outputs [86]. After model fitting, we computed estimates of marginal means of the posterior distributions of terms of interest with their two-tailed posterior probabilities (PPs) against zero as $PP = 2 \times \min [P(x > 0), P(x < 0)]$. For estimating marginal means, we used the emmeans (<https://github.com/rvnlenth/emmeans>) R package.

Posterior predictive checks, combined with visual comparisons between simulated and observed data, confirmed that the models accurately captured the underlying data structure and met the necessary assumptions. These checks also ensured model convergence. Convergence was further verified by evaluating potential scale reduction factors and effective bulk and tail sample sizes. The use of weakly informative zero-mean normal priors, model selection through LOO, and testing multiple distributions with monotonic effects for the variables helped capture non-linear relationships and ensured reliable parameter estimation and model robustness.

5. Results

The following sections present the results of each analysis.

5.1. Factor Constructs for Trust, Benefit, Intention, and Interaction Depth

We fitted a five-factor CFA model with 20 items for our data with 154 samples. Out of these 154, 38 subjects were missing one or two values, which were handled by the ML algorithm of Lavaan. Out of the original 24 items, 4 were left out from the ATI group (questions 3, 6, 7, and 8) due to poor fit in any construct and also to keep the number of items per factor at a maximum of five. If included, those four items had standardized weights less than 0.6 and were found to decrease the fit indices of the model. Factors, items, and fit parameters of the final five-factor model with 20 items are listed in Table 2.

Table 2. Parameters of the five-factor confirmatory factor model, including standardized estimates, Z-values, and confidence intervals (CIs). The model was fitted with $n = 154$ samples.

Factor Construct	Item	Est. Std	Z-Val	CI. Lower	CI. Upper
Behavioral intention to use AI-based educational tools	Plan to use	0.914	35.358	0.863	0.964
	Possibility to use	0.648	12.463	0.546	0.750
	Use if opportunity	0.867	29.399	0.809	0.925
Perceived benefits	Enhance learning	0.769	19.671	0.693	0.846
	Improved studies	0.813	23.857	0.746	0.880
	Increased productivity	0.779	20.420	0.704	0.854
	Accelerated studies	0.644	12.194	0.541	0.748
Trust in AI-based essay grading tool	Found useful	0.803	22.714	0.734	0.872
	Believable	0.867	33.661	0.816	0.917
	Accurate	0.859	31.780	0.806	0.912
	Confidence	0.866	32.836	0.814	0.917
Trust in AI applications	Reliable	0.812	24.974	0.748	0.876
	Easy to trust	0.885	35.182	0.836	0.934
	High tendency	0.971	50.471	0.934	1.009
Affinity for technology interaction	Blindly trusting	0.748	19.138	0.671	0.824
	Detail focus	0.733	17.022	0.649	0.817
	Test functions	0.843	27.871	0.783	0.902
	Intensive trial	0.830	26.949	0.770	0.890
	Enjoy learning	0.885	35.949	0.837	0.933
	Full use	0.713	15.792	0.624	0.801

Reliability estimated for factors are listed in Table 3.

Table 3. The reliability measures Alpha, Omega, and average variance explained (*Avevar*) of the five-factor confirmatory factor model.

Measure	Behavioral Intention	Perceived Benefits	Trust in AI-Based Essay Grading Tool	Trust in AI Applications	Affinity for Technology Interaction
Alpha	0.8498	0.8729	0.9119	0.9002	0.8981
Omega	0.8644	0.8716	0.9128	0.9007	0.9018
Avevar	0.6878	0.5770	0.7245	0.7529	0.6506

For this model, the Comparative Fit Index (CFI) was 0.951, the Tucker–Lewis Index (TLI) was 0.942, and the RMSEA was 0.064. The CFA results demonstrate that the five-factor model provides a good fit to the data, as indicated by the fit indices (CFI and TLI > 0.90, RMSEA < 0.08). The standardized coefficients for the items associated with each construct are significant and demonstrate substantial loadings on their respective factors, supporting the constructs’ validity. The reliability measures (Alpha and Omega) for each construct suggest a high level of internal consistency, confirming the reliability of the constructs for further analysis.

All pair-wise covariances (correlations) between standardized factors are listed in Table 4. The result indicates that there is a notable correlation ($p < 0.05$) between all but three pairs (perceived benefits—affinity for technology interaction; trust in AI-based essay

grading tool—affinity for technology interaction; trust in AI applications—affinity for technology interaction).

Table 4. Estimated correlations between the five factors.

Factor 1	Factor 2	Estimate	z-Value	p (> z)
Behavioral intention to use AI-based educational tools	Perceived benefits	0.7350	15.2990	0
	Trust in AI-based essay grading tool	0.3720	4.7170	0
	Trust in AI applications	0.5270	7.9870	0
	Affinity for technology interaction	0.1990	2.2910	0.0220
Perceived benefits	Trust in AI-based essay grading tool	0.6050	9.8260	0
	Trust in AI applications	0.4700	6.6040	0
	Affinity for technology interaction	0.1220	1.3710	0.1700
Trust in AI-based essay grading tool	Trust in AI applications	0.5100	7.7460	0
	Affinity for technology interaction	0.0730	0.8310	0.4060
Trust in AI applications	Affinity for technology interaction	0.1470	1.7390	0.0820

From these correlations, we can conclude that students who had higher trust in AI applications also had higher trust in AI-based essay grading tools (correlation 0.51), perceived benefits (0.47), and intention to use AI-based educational tools (0.527) with all estimated correlation non-zero with $p < 0.0005$. These outcomes were also verified by regression models with other variables present (see Tables 6, 8 and 10). Hence, we can confirm Hypotheses 15–17.

5.2. Regression Analysis

Using the five-factor model, we fitted all four Bayesian regression models using those samples with all variables present in the demographics group (n = 142), i.e., we dropped 12 subjects compared to our factor analysis.

5.2.1. Preference of AI over Teacher

The suitable model supported by the data was as follows:

$$\text{preferring AI over teacher} \sim 1 + \text{age group} + \text{trust in AI application} \times \text{affinity for technology interaction}$$

with Bayesian $R^2 = 0.09067$ (HDI 95% 0.0249–0.1715) assuming linear scale. Fitted model coefficients are shown in Table 5. Only the linear trend for Trust in AI applications was found to be non-zero 0.4357 at $p < 0.01$.

Table 5. The fit coefficient for the regression model of estimating variable preferring AI over teacher, including HDI at 95%. Intercept and simplex terms are omitted.

Term	Estimate	Est. Error	Q2.5	Q97.5
Trust in AI applications	0.4368	0.1715	0.1017	0.7745
Affinity for technology interaction	0.1000	0.1650	−0.2193	0.4247
Trust in AI applications: affinity for technology interaction	0.2015	0.1505	−0.0965	0.4932
Age group	0.1762	0.2633	−0.3445	0.7077

Estimated marginal means for age group levels were 1 = 0 (fixed reference), 2 = 0.071526 (−0.50752–0.730111), 3 = 0.262692 (−0.68016–1.244505), and 4 = 0.52364 (−1.0607–2.084086)

using HDI 95% (in parenthesis). No notable differences were found between any age group pairs (i.e., all $PP > 0.05$ for including zero).

5.2.2. Benefits of Using an AI-Based Essay Grading Tool

The suitable model supported by the data was as follows:

$$\text{perceived benefits} \sim 1 + \text{education program} + \text{preferred study style} \times \text{preferred study location} + \text{preferring AI over teacher} + \text{trust in AI applications} + \text{affinity for technology interaction}$$

with Bayesian R^2 of 0.371 (HDI 95% 0.265–0.461). Fitted model coefficients are shown in Table 6. Non-zero trends ($PP < 0.05$) were found for the covariates trust in AI applications and preferring AI over teachers.

Table 6. The fit coefficient for the regression model of estimating variable perceived benefits, including HDI at 95%. Intercept and simplex terms are omitted.

Term	Estimate	Est. Error	Q2.5	Q97.5
Education program 2	−0.2828	0.1645	−0.6077	0.0457
Education program 4	−0.7954	0.4317	−1.6359	0.0509
Preferred study style 2	0.2656	0.3236	−0.3687	0.9019
Preferred study location 2	−0.1384	0.1670	−0.4680	0.1848
Trust in AI applications	0.3788	0.0742	0.2343	0.5244
Affinity for technology interaction	0.0556	0.0721	−0.0857	0.1963
Preferred study style 2: preferred study location 2	−0.1124	0.3705	−0.8422	0.6179
Preferring AI over teacher	0.3132	0.0990	0.1365	0.5269

For the education program, marginal mean estimates were 1 = 0.333214 (HDI 95% −0.00572–0.676204), 2 = 0.049095 (−0.17096–0.273835), and 4 = −0.46063 (−1.27562–0.371828). No notable differences were found between different programs (zero included for all at $PP < 0.05$).

Marginal mean estimates for preferring AI over teacher, preferred study style, and preferred study location are visualized in Figure 2.

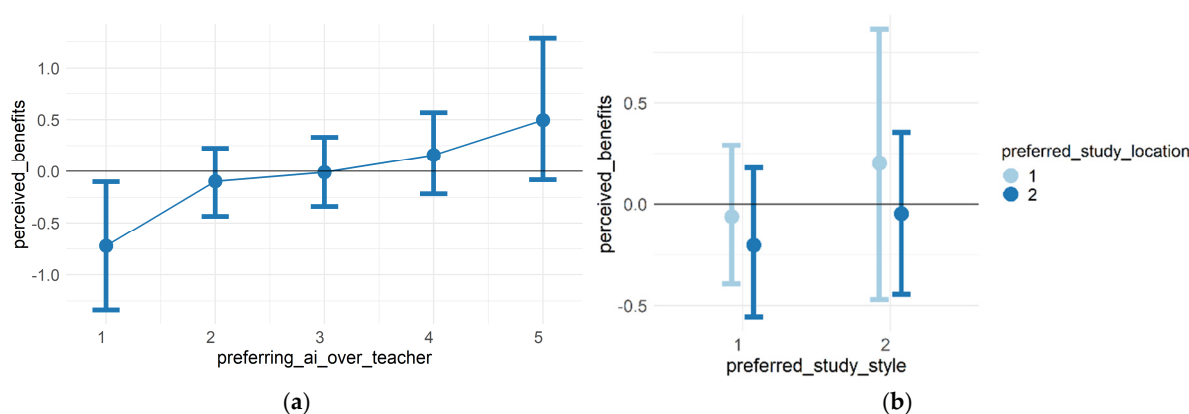


Figure 2. Marginal means for variables (a) preferring AI over teacher, and (b) preferred study style and preferred study location. Error bars indicate HDI 95%.

Differences between levels of preferring AI over teacher were all non-zero at $PP < 0.05$ for all pairs except between levels 4 and 5. For preferred study style and preferred study

location, no notable differences were found (zero was included with $PP < 0.05$) between any pairs, including six mixed pairs.

With these results, we can respond to the hypotheses about the perceived benefits of AI-based essay grading. The results are presented in Table 7.

Table 7. The results of tested hypotheses about perceived benefits of AI-based essay grading.

H3. <i>There is a significant association between students' demographics and perceived benefits of AI-based essay grading.</i>	Not supported. Our data did not support including demographic variables in the model.
H7. <i>Students who prefer to study alone perceive more benefits from AI-based essay grading</i>	Inconclusive. The perceived benefit was lower for those preferring studying alone; however, we could not rule out zero ($PP > 0.05$) between differences.
H10. <i>Students who believe that AI-based assessment is more effective than teacher feedback perceive more benefits from AI-based essay grading</i>	Supported. Students who believe that AI-based assessment is more effective than teacher feedback perceive more benefits from AI-based essay grading.
H13. <i>Students' affinity for technology interaction does not significantly influence their perceived benefits of AI-based essay grading.</i>	Supported. We found no notable trend (0.0556) between affinity for technology interaction and perceived benefits of AI-based essay grading.

5.2.3. Trust for AI-Based Essay Grading Tool

The suitable model supported by the data was as follows:

$$\text{trust in AI-based essay grading tool} \sim 1 + \text{preferring AI over teacher} + \text{trust in AI applications} + \text{affinity for technology interaction}$$

with Bayesian R^2 of 0.3108741 (0.2044818–0.4136357). Fitted model coefficients are shown in Table 8. Non-zero trends were found for covariates' trust in AI applications and preferring AI over teacher.

Table 8. The fit coefficient for the regression model of estimating the variable trust in AI-based essay grading tool, including HDI at 95%. Intercept and simplex terms are omitted.

Term	Estimate	Est. Error	Q2.5	Q97.5
Trust in AI applications	0.3834	0.0638	0.2597	0.5108
Affinity for technology interaction	0.0170	0.0616	−0.105	0.1366
Preferring AI over teacher	0.2461	0.0774	0.1053	0.4125

Marginal mean estimates for preferring AI over teacher are visualized in Figure 3. All pair-wise comparisons between levels were non-zero at level $PP < 0.05$.

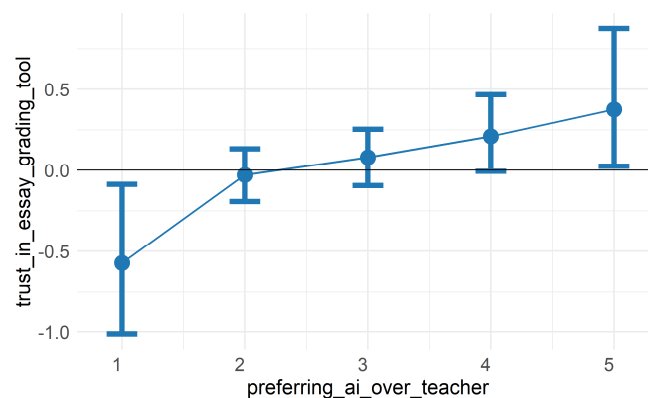


Figure 3. Marginal means for variable preferring AI over teacher. Error bars indicate HDI 95%.

With these results, we can respond to hypotheses about perceived trust in AI-based essay grading. The results are presented in Table 9.

Table 9. The results of tested hypotheses about perceived trust in AI-based essay grading.

H1. Younger students (under 40) have higher trust in AI-based essay grading than older students	Not supported. Our data did not support including demographic variables in the model.
H2. Male students tend to show more trust in AI-based essay grading compared to other genders	Not supported. Our data did not support including demographic variables in the model.
H5. Students of IT or other digital education programs have higher trust in the AI-based essay grading tool than the students of other degree programs.	Not supported. Our data did not support including program variables in the model.
H8. Students who prefer to study alone have more trust in AI-based essay grading.	Not supported. Our data did not support including the study style variable in the model.
H12. Student’s affinity for technology interaction does not significantly influence their trust in AI-based essay grading.	Supported. We found no notable trend (0.0170) between affinity for technology interaction and perceived benefits of AI-based essay grading.
H9. Students who believe that AI-based assessment is more effective than teacher feedback have more trust in AI-based essay grading	Supported. Students who believed that AI-based assessment is more effective than teacher feedback found the tool notably more trustful (trend 0.2461).

5.2.4. Behavioral Intention to Use AI Tools in Education

The suitable model supported by the data was as follows:

$$\text{behavioral intention to use AI-based educational tools} \sim 1 + \text{education program} + \text{preferred study style} \times \text{preferred study location} + \text{preferring AI over teacher} + \text{trust in AI applications} + \text{affinity for technology interaction}$$

with Bayesian R^2 of 0.3653 (0.2580–0.4606). Fitted model coefficients are shown in Table 10. Non-zero trends were found for covariates’ trust in AI applications and preferring AI over teacher.

Table 10. The fit coefficient for the regression model of estimating variable behavioral intention to use AI-based educational tools, including HDI at 95%. Intercept and simplex terms are omitted.

Term	Estimate	Est. Error	Q2.5	Q97.5
Education program 2	−0.2776	0.1515	−0.5821	0.0132
Education program 4	−0.6489	0.4126	−1.4085	0.2095
Preferred study style 2	0.3026	0.2947	−0.2493	0.8974
Preferred study location 2	−0.2992	0.1564	−0.6014	0.0070
Trust in AI applications	0.4485	0.0684	0.3184	0.5881
Affinity for technology interaction	0.0999	0.0643	−0.0289	0.2230
Preferred study style 2: preferred study location 2	−0.3147	0.3363	−0.9873	0.3290
Preferring AI over teacher	0.1421	0.0861	−0.0107	0.3253

Estimated marginal means for the education program were 1 = 0.319788 (0.001414–0.629917), 2 = 0.042989 (−0.15–0.247714), and 4 = −0.34609 (−1.07781–0.474569). All three pair-wise differences between these were not different from zero ($PP < 0.05$).

Marginal mean estimates for preferring AI over teacher, preferred study style, and preferred study location are visualized in Figure 4.

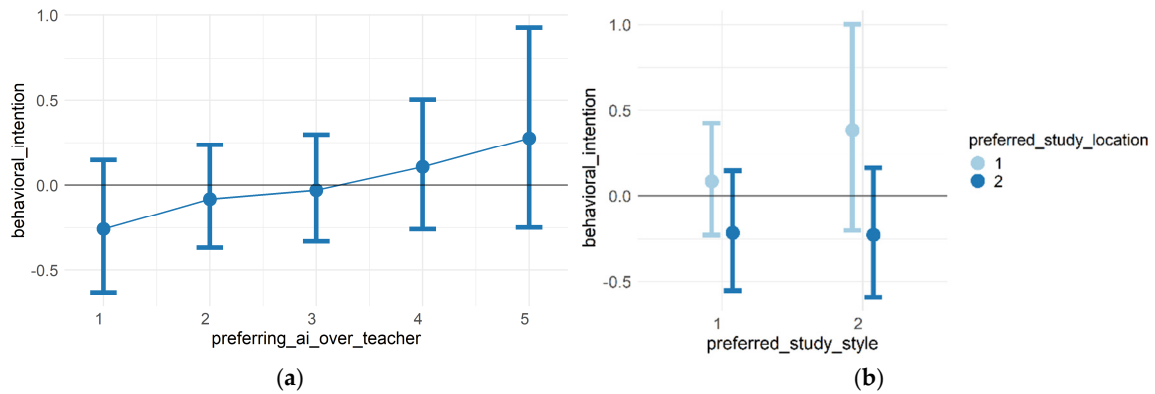


Figure 4. Marginal means for variables (a) preferring AI over teacher, and (b) preferred study style and preferred study location. Error bars indicate HDI 95%.

None of the pair-wise differences for preferring AI over teacher were non-zero at the level of $PP < 0.05$. For preferred study style and preferred study location, we found three notable non-zero differences listed in Table 11.

Table 11. Estimated contrasts for behavioral intention to use AI tools involving variables preferred study style and preferred study location were found notable non-zero. * = $PP < 0.05$, ** = $PP < 0.01$.

Contrast	Estimate	Lower. HP	Upper. HP	
Preferred study style = 2 and preferred study location = 1 - preferred study style = 1 and preferred study location = 2	0.5920	0.0183	1.1946	*
Preferred study style = 2 and preferred study location = 1 - preferred study style = 2 and preferred study location = 2	0.6068	0.0435	1.2294	*
Preferred study location = 1 - preferred study location = 2	0.4543	0.0125	0.9100	**

With these results, we can respond to the hypotheses about the intention to use AI-based tools in the future. The results are presented in Table 12.

Table 12. The results of tested hypotheses about the intention to use AI-based tools in the future.

H4. There is a significant association between students' demographics and behavioral intention to use AI-based educational tools in the future.	Not supported. Our data did not support including demographic variables in the model.
H6. Students who prefer to study remotely have a higher behavioral intention to use AI-based educational tools	Supported. Intention to use AI-based educational tools was higher for those who prefer to study remotely (difference 0.45).
H11. Students who believe that AI-based assessment is less effective than teacher feedback have a lower behavioral intention to use AI-based educational tools.	Inconclusive. We noticed a positive trend (0.1421) but could not rule out zero ($PP > 0.05$).
H14. Students with a higher affinity for technology interaction have higher behavioral intention to use AI-based educational tools.	Inconclusive. Although there was a clear positive trend, it was not notably different from zero.

6. Discussion

In this study, students used an AI-based essay scoring system in a real learning situation, and data were collected in this context. The study provides a novel contribution to understanding the relationship between learning preferences and AI-based assessment and essay scoring. Prior studies between learning preferences and trust in AI-based grading are scant, although the studies on AI-based applications in online learning show an increasing trend [39].

Trust in AI applications and AI-based essay grading tools is important in shaping users' perceived benefits and their behavioral intention to adopt these technologies. Trust

builds confidence in the AI's reliability, fairness, and accuracy, which, in turn, enhances users' perception of the utility and value these tools provide. When users see tangible benefits, such as improved grading consistency and personalized feedback, their intention to continue using these AI tools in educational contexts strengthens. The interplay between trust, perceived benefits, and behavioral intention highlights a cycle where trust fosters perceived value, and perceived value further reinforces user trust, ultimately promoting adoption and sustained engagement with AI technologies in education.

The key finding of this study is that the main driving factor behind AI-based assessment over teacher, trust in AI-based essay grading, perceived benefits, and usage intention was general trust in AI applications with positive trends of 0.44, 0.38, 0.38, and 0.45, respectively—all significant with $PP < 0.05$. Thus, general trust in AI applications is a strong predictor for AI education solutions. This supports the finding of Roy et al. (2022) [59], who stated that trust is an essential factor for AI adoption in education. Trust is an important concept to study in education, as it affects the extent to which individuals believe technology will create value and solve their needs in a secure and trustworthy way. Prior research shows that students may also feel anxious and less confident when learning with AI [1]. Regarding demographic information and students' preferences, no significant correlation was found with trust in the AI-based essay grading tool, perceived benefits, or behavioral intention to use such a system. These results suggest that age, gender, study preferences, and the educational program do not notably influence students' trust or their perception of the benefits of AI-based essay grading. The lack of support for these hypotheses may imply that other factors could be more influential in shaping students' experiences with and attitudes toward AI technology in education. For instance, previous exposure to technology and personal interest may play crucial roles.

Prior studies [47,87] have found that AI-based essay scoring is not valued more than teachers' assessments. Our study found that this should not be generalized, as students' learning preferences affect their selection. We found that learning preference was a factor that influenced the intention to use AI-based essay grading. That is, our study found that students who favor remote learning show a greater inclination to utilize AI-based educational tools. The perceived benefits of AI-based essay grading among those who prefer to study alone, however, remain inconclusive. We observed that students who rate the effectiveness of AI-based assessment above teacher feedback place more trust in such tools and perceive more benefits from them. However, it remains inconclusive as to whether this belief also leads to a stronger intention to make use of AI-based educational tools in the future.

Our study found no statistically significant relationship (only a weaker positive association) between affinity for technology interaction (ATI) and trust in AI-based essay grading tools. This aligns with the findings by Kahr et al. (2023) [53] and Schadelbauer et al. (2023) [54], who observed similar positive but non-significant associations. However, it contrasts with Brauner et al. (2023) [52], who reported a significant positive relationship between ATI and trust in AI applications. However, Brauner et al. (2023) [52] also noted that user evaluations and expectations of AI vary across domains, with users often struggling to assess AI's opportunities and risks. In our educational context, students' ATI did not significantly influence their trust in AI-based essay grading or their intention to use such tools. This discrepancy may stem from the specific nature of educational AI applications, where trust is influenced more by perceived fairness and transparency than by general technology affinity. Further research is needed to explore these domain-specific trust factors.

Although the ATI results are not statistically significant, our analysis still showed a positive relationship between ATI and affection towards AI-based essay grading systems. Namely, the results of the ATI scale indicate positive trends for all four responses (estimates 0.10, 0.06, 0.02, and 0.10); however, none of these correlations are strong enough ($PP > 0.05$). Hence, we conclude that students' affinity for technology interaction does not significantly influence their trust in AI-based essay grading, perceived benefits of AI-based essay grading,

and behavioral intention to use AI-based educational tools. The results of the ATI scales are not directly related to AI or other educational technologies; hence, they cannot be considered an explaining factor of how students perceive the usage of AI in education or AI-based essay assessment. We conclude that students' interest in technological systems in general does not directly indicate interest in AI or other education technologies.

We further conclude that our experiment settings might have affected the results. We studied students' ATI, trust, behavioral intention to use AI-based educational tools, and preferences in an authentic learning and teaching situation where students used an AI-based educational solution in a real-life context. We expected the students to reflect on their usage experiences and preferences from the viewpoint of learning and teaching, not from the viewpoint of using a new technology. The AI-based essay grading application was an enabler for automatic and immediate assessment, not the primary focus of a learning task. This might explain why the ATI scales did not predict trust and behavioral intention towards AI-based educational solutions. Students do not define their preferences towards AI-based assessment solutions based on their personal affective orientation towards new technologies. Unlike affinity for technology interaction in general, learning preferences predicted the intention to use AI in education. Learning preferences are directly linked to learning and teaching in both AI-based automatic assessment and teacher-led assessment situations. Hence, we expect that students evaluate new AI-based assessment solutions from the viewpoint of their learning preferences rather than their orientation towards new technologies.

In summary, this study contributes to a better understanding of how students perceive AI-assisted assessments in higher education contexts while opening avenues for further explorations in this emerging field. Achieving a deeper understanding of these issues will allow us to design and implement more effective AI-based educational tools that better align with students' needs and preferences.

This study has some limitations that must be acknowledged. The basic limitations of this study include the relatively focused target group of information technology and business students in higher education, a short usage experience consisting of one learning assignment evaluated by AI, and a quite simple assessment process focusing on the generation of correct factual elements. The sample size, while sufficient for the confirmatory factor analysis and Bayesian regression, may limit the generalizability of the findings to broader student populations. Future research with a larger and more diverse sample, including students from various cultural and institutional backgrounds, is necessary to validate these results and account for potential variability in trust and behavioral intentions across different demographic groups.

The study's reliance on a single AI-based essay grading tool introduces potential biases related to the tool's specific features, such as its training data and feedback mechanisms. While efforts were made to minimize these biases during the training phase, the findings may not be fully generalizable to other AI systems with different architectures or functionalities. Future studies should replicate this research using alternative AI-based tools to ensure consistency and broader applicability.

While we noticed that 50–100-word responses were enough to assess student perceptions using short essay responses (50–100 words), this length may still not have captured the full range of students' engagement and trust in the AI tool. Longer, more complex essays might reveal different dynamics in trust, perceived benefits, and behavioral intentions.

Additionally, while demographic and study preference variables were included, other potential moderating factors, such as students' prior experience with AI technologies, individual levels of AI literacy, or disciplinary focus (e.g., STEM vs. non-STEM fields), were not deeply investigated. These variables may play an important role in shaping attitudes toward AI-based tools and should be explored in future studies.

Finally, the cross-sectional design of the study limits our ability to infer causality between the variables examined. Longitudinal studies could provide a more comprehensive

understanding of how trust, perceived benefits, and behavioral intentions evolve over time with increased exposure to and familiarity with AI-based educational tools.

7. Conclusions and Future Work

This study contributes to the discussion of AI in higher education and, more specifically, to AI-based formative assessment. We found that general trust in AI applications predicts the use of other AI applications in education. In addition, students' learning preferences, like preferring to study alone and remotely, affect their intention to use AI-based essay scoring systems. However, a positive affinity for technology interaction does not explain perceived benefits, trust, and intention to use AI-based essay grading. Thus, we conclude that students prefer AI-based essay assessment based on their learning orientation, not based on their demographics or technological viewpoint. Overall, our study underlines the importance of students' learning preferences, as not all students are similar in their learning orientation.

This study also encourages researchers to examine students' learning orientation and preferences regarding AI-based learning and teaching since effective instruction and usage of AI tools should be student-centric and adaptive to various learning needs and priorities. Virtual learning, MOOCs (massive open online courses), and distance learning are continuously growing, especially in higher education, and there is a clear need for new knowledge regarding AI-based assessment and scoring. Thus, the ever-changing relationship of students, teachers, and AI technologies calls for more research.

In future work, it would thus be beneficial to conduct similar studies with more diverse samples to confirm whether our findings are representative of larger populations. It may also be valuable to explore other factors, such as prior technology experience or personal interest, which might play crucial roles in shaping attitudes towards AI technologies in education. Moreover, longitudinal studies could provide more insights into how these attitudes evolve over time as users become more familiar with the system. Furthermore, as AI continues to advance and its use becomes more pervasive in educational settings, ongoing assessment of its acceptance among users/students will remain an important area for future research.

Author Contributions: Conceptualization, U.A.K.; Methodology, U.A.K. and J.K.; Software, J.K.; Formal analysis, U.A.K. and J.K.; Data curation, A.A.; Writing—original draft, A.A., U.A.K., J.K. and S.S.; Writing—review and editing, A.A., U.A.K. and S.S.; Visualization, J.K.; Funding acquisition, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: Ministry of Education and Culture, Finland, OKM/108/523/2021.

Institutional Review Board Statement: MCI Ethics Committee 20221203, approved on 15 December 2022.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chiu, T.K.F.; Xia, Q.; Zhou, X.; Chai, C.S.; Cheng, M. Systematic Literature Review on Opportunities, Challenges, and Future Research Recommendations of Artificial Intelligence in Education. *Comput. Educ. Artif. Intell.* **2023**, *4*, 100118. [[CrossRef](#)]
2. Zhai, X.; Chu, X.; Chai, C.S.; Jong, M.S.Y.; Istenic, A.; Spector, M.; Liu, J.-B.; Yuan, J.; Li, Y. A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity* **2021**, *2021*, 8812542. [[CrossRef](#)]
3. Nemorin, S.; Vlachidis, A.; Ayerakwa, H.M.; Andriotis, P. AI Hyped? A Horizon Scan of Discourse on Artificial Intelligence in Education (AIED) and Development. *Learn. Media Technol.* **2023**, *48*, 38–51. [[CrossRef](#)]
4. Delcker, J.; Heil, J.; Ifenthaler, D.; Seufert, S.; Spirgi, L. First-Year Students' AI Competence as a Predictor for Intended and de Facto Use of AI Tools for Supporting Learning Processes in Higher Education. *Int. J. Educ. Technol. High. Educ.* **2024**, *21*, 18. [[CrossRef](#)]

5. Gao, R.; Merzdorf, H.E.; Anwar, S.; Hipwell, M.C.; Srinivasa, A. Automatic Assessment of Text-Based Responses in Post-Secondary Education: A Systematic Review. *Comput. Educ. Artif. Intell.* **2024**, *6*, 100206. [CrossRef]
6. Ramesh, D.; Sanampudi, S.K. An Automated Essay Scoring Systems: A Systematic Literature Review. *Artif. Intell. Rev.* **2022**, *55*, 2495–2527. [CrossRef]
7. Minn, S. AI-Assisted Knowledge Assessment Techniques for Adaptive Learning Environments. *Comput. Educ. Artif. Intell.* **2022**, *3*, 100050. [CrossRef]
8. Mononen, A.; Alamäki, A.; Kauttonen, J.; Klemetti, A.; Passi-Rauste, A.; Ketamo, H. Forecasted Self: AI-Based Careerbot-Service Helping Students with Job Market Dynamics. *Eng. Proc.* **2023**, *39*, 99. [CrossRef]
9. Stanja, J.; Gritz, W.; Krugel, J.; Hoppe, A.; Dannemann, S. Formative Assessment Strategies for Students' Conceptions—The Potential of Learning Analytics. *Br. J. Educ. Technol.* **2023**, *54*, 58–75. [CrossRef]
10. Chai, C.S.; Wang, X.; Xu, C. An Extended Theory of Planned Behavior for the Modelling of Chinese Secondary School Students' Intention to Learn Artificial Intelligence. *Mathematics* **2020**, *8*, 2089. [CrossRef]
11. Bakhadirov, M.; Alasgarova, R. Factors Influencing Teachers' Use of Artificial Intelligence for Instructional Purposes. *IAFOR J. Educ.* **2024**, *12*, 9–32. [CrossRef]
12. Viberg, O.; Cukurova, M.; Feldman-Maggor, Y.; Alexandron, G.; Shirai, S.; Kanemune, S.; Wasson, B.; Tømte, C.; Spikol, D.; Milrad, M.; et al. What Explains Teachers' Trust in AI in Education across Six Countries? *Int. J. Artif. Intell. Educ.* **2024**. [CrossRef]
13. Almaiah, M.A.; Alfaisal, R.; Salloum, S.A.; Hajjeh, F.; Thabit, S.; El-Qirem, F.A.; Lutfi, A.; Alrawad, M.; Al Mulhem, A.; Alkhdour, T.; et al. Examining the Impact of Artificial Intelligence and Social and Computer Anxiety in E-Learning Settings: Students' Perceptions at the University Level. *Electronics* **2022**, *11*, 3662. [CrossRef]
14. Ke, Z.; Inamdar, H.; Lin, H.; Ng, V. Give Me More Feedback II: Annotating Thesis Strength and Related Attributes in Student Essays. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 3994–4004.
15. Gikandi, J.W.; Morrow, D.; Davis, N.E. Online Formative Assessment in Higher Education: A Review of the Literature. *Comput. Educ.* **2011**, *57*, 2333–2351. [CrossRef]
16. Wiliam, D.; Black, P. Meanings and Consequences: A Basis for Distinguishing Formative and Summative Functions of Assessment? *Br. Educ. Res. J.* **1996**, *22*, 537–548. [CrossRef]
17. Dirin, A.; Alamäki, A.; Suomala, J. Gender Differences in Perceptions of Conventional Video, Virtual Reality and Augmented Reality. 2019. Available online: <https://www.learntechlib.org/p/216491/> (accessed on 11 December 2024).
18. Yang, A.C.M.; Chen, I.Y.L.; Flanagan, B.; Ogata, H. How Students' Self-Assessment Behavior Affects Their Online Learning Performance. *Comput. Educ. Artif. Intell.* **2022**, *3*, 100058. [CrossRef]
19. Clark, R.E. Media Will Never Influence Learning. *Educ. Technol. Res. Dev.* **1994**, *42*, 21–29. [CrossRef]
20. Alamäki, A.; Nyberg, C.; Kimberley, A.; Salonen, A.O. Artificial Intelligence Literacy in Sustainable Development: A Learning Experiment in Higher Education. *Front. Educ.* **2024**, *9*, 1343406. [CrossRef]
21. Kamii, C. The Equilibration of Cognitive Structures: The Central Problem of Intellectual Development. *Am. J. Educ.* **1986**, *94*, 574–577. [CrossRef]
22. Vygotsky, L.S.; Cole, M. *Mind in Society: Development of Higher Psychological Processes*; Harvard University Press: Cambridge, MA, USA, 1978.
23. Kaya, F.; Aydin, F.; Schepman, A.; Rodway, P.; Yetişensoy, O.; Demir Kaya, M. The Roles of Personality Traits, AI Anxiety, and Demographic Factors in Attitudes toward Artificial Intelligence. *Int. J. Hum. Comput. Interact.* **2024**, *40*, 497–514. [CrossRef]
24. Rogers, E.M. *Diffusion of Innovations*, 5th ed.; Free Press: New York, NY, USA, 2003.
25. Dunnhumby Understanding Consumer Trust in Artificial Intelligence 2023. Available online: <https://www.dunnhumby.com/> (accessed on 11 December 2024).
26. Gillespie, N.; Lockey, S.; Curtis, C.; Pool, J.; Akbari, A. *Trust in Artificial Intelligence: A Global Study*; The University of Queensland: Brisbane, Australia; KPMG Australia: Sydney, Australia, 2023.
27. Bem, S.L. Gender Schema Theory: A Cognitive Account of Sex Typing. *Psychol. Rev.* **1981**, *88*, 354–364. [CrossRef]
28. Lebow, S. Men More Likely than Women to Trust Generative AI. *Morning Consult.* 2023. Available online: <https://www.insiderintelligence.com/content/men-more-likely-than-women-trust-generative-ai> (accessed on 11 December 2024).
29. Mearian, L. With AI, There's a Trust Gap Based on Gender, Age. *Computerworld.* 2023. Available online: <https://www.computerworld.com/article/3707795/ai-trust-gap-based-on-gender-age.html> (accessed on 11 December 2024).
30. Gefen, D.; Straub, D.W. Gender Differences in the Perception and Use of E-Mail: An Extension to the Technology Acceptance Model. *MIS Q.* **1997**, *21*, 389–400. [CrossRef]
31. McKnight, D.H.; Carter, M.; Thatcher, J.B.; Clay, P.F. Trust in a Specific Technology: An Investigation of Its Components and Measures. *ACM Trans. Manag. Inf. Syst. TMIS* **2011**, *2*, 1–25. [CrossRef]
32. Omrani, N.; Riviuccio, G.; Fiore, U.; Schiavone, F.; Garcia Agreda, S. To Trust or Not to Trust? An Assessment of Trust in AI-Based Systems: Concerns, Ethics and Contexts. *Technol. Forecast. Soc. Chang.* **2022**, *181*, 121763. [CrossRef]
33. Castillo-Acobo, R.; Tiza, D.; Orellana, M.; Cajigas, B.; Huayta-Meza, F.; Sota, C.; Muñoz, I.; Acevedo, J.; Sernaqué, A.; Carranza, C.; et al. Artificial Intelligence Application in Education. *J. Namib. Stud. Hist. Politics Cult.* **2023**, *33*, 792–807. [CrossRef]
34. Davis, F.D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **1989**, *13*, 319–340. [CrossRef]

35. Hornberger, M.; Bowersdorff, A.; Nerdel, C. What Do University Students Know about Artificial Intelligence? Development and Validation of an AI Literacy Test. *Comput. Educ. Artif. Intell.* **2023**, *5*, 100165. [CrossRef]
36. Marrone, R.; Taddeo, V.; Hill, G. Creativity and Artificial Intelligence—A Student Perspective. *J. Intell.* **2022**, *10*, 65. [CrossRef] [PubMed]
37. Long, D.; Magerko, B. What Is AI Literacy? Competencies and Design Considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–16.
38. Venkatesh, V.; Morris, M.G.; Davis, G.B.; Davis, F.D. User Acceptance of Information Technology: Toward a Unified View. *MIS Q.* **2003**, *27*, 425–478. [CrossRef]
39. Jia, K.; Wang, P.; Li, Y.; Chen, Z.; Jiang, X.; Lin, C.-L.; Chin, T. Research Landscape of Artificial Intelligence and E-Learning: A Bibliometric Research. *Front. Psychol.* **2022**, *13*, 795039. [CrossRef] [PubMed]
40. Seo, K.; Tang, J.; Roll, I.; Fels, S.; Yoon, D. The Impact of Artificial Intelligence on Learner–Instructor Interaction in Online Learning. *Int. J. Educ. Technol. High. Educ.* **2021**, *18*, 54. [CrossRef] [PubMed]
41. Jin, S.-H.; Im, K.; Yoo, M.; Roll, I.; Seo, K. Supporting Students’ Self-Regulated Learning in Online Learning Using Artificial Intelligence Applications. *Int. J. Educ. Technol. High. Educ.* **2023**, *20*, 37. [CrossRef]
42. Zimmerman, B.J. Becoming a Self-Regulated Learner: An Overview. *Theory Pract.* **2002**, *41*, 64–70. [CrossRef]
43. Wei, X.; Sun, S.; Wu, D.; Zhou, L. Personalized Online Learning Resource Recommendation Based on Artificial Intelligence and Educational Psychology. *Front. Psychol.* **2021**, *12*, 767837. [CrossRef] [PubMed]
44. Colonna, L. Teachers in the Loop? An Analysis of Automatic Assessment Systems under Article 22 GDPR. *Int. Data Priv. Law* **2023**, *14*, 3–18. [CrossRef]
45. Escalante, J.; Pack, A.; Barrett, A. AI-Generated Feedback on Writing: Insights into Efficacy and ENL Student Preference. *Int. J. Educ. Technol. High. Educ.* **2023**, *20*, 57. [CrossRef]
46. Celik, I.; Dindar, M.; Muukkonen, H.; Järvelä, S. The Promises and Challenges of Artificial Intelligence for Teachers: A Systematic Review of Research. *TechTrends* **2022**, *66*, 616–630. [CrossRef]
47. Stoica, E. A Student’s Take on Challenges of AI-Driven Grading in Higher Education. Bachelor’s Thesis, University of Twente, Enschede, The Netherlands, 2022.
48. Nazaretsky, T.; Cukurova, M.; Alexandron, G. An Instrument for Measuring Teachers’ Trust in AI-Based Educational Technology. In Proceedings of the LAK22: 12th International Learning Analytics and Knowledge Conference, Online, 21–25 March 2021; pp. 1–6.
49. Nazaretsky, T.; Ariely, M.; Cukurova, M.; Alexandron, G. Teachers’ Trust in AI-Powered Educational Technology and a Professional Development Program to Improve It. *Br. J. Educ. Technol.* **2022**, *53*, 914–931. [CrossRef]
50. Nazaretsky, T.; Cukurova, M.; Ariely, M.; Alexandron, G. Confirmation Bias and Trust: Human Factors That Influence Teachers’ Attitudes Towards AI-Based Educational Technology. In Proceedings of the CEUR Workshop Proceedings, Uzhhorod, Ukraine, 28–30 September 2021; Volume 3042.
51. Franke, T.; Attig, C.; Wessel, D. A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *Int. J. Hum. Comput. Interact.* **2019**, *35*, 456–467. [CrossRef]
52. Brauner, P.; Hick, A.; Philipsen, R.; Ziefle, M. What Does the Public Think about Artificial Intelligence?—A Criticality Map to Understand Bias in the Public Perception of AI. *Front. Comput. Sci.* **2023**, *5*, 1113903. [CrossRef]
53. Kahr, P.K.; Rooks, G.; Willemsen, M.C.; Sniijders, C.C.P. It Seems Smart, but It Acts Stupid: Development of Trust in AI Advice in a Repeated Legal Decision-Making Task. In Proceedings of the 28th International Conference on Intelligent User Interfaces, Sydney, NSW, Australia, 27–31 March 2023; pp. 528–539.
54. Schadelbauer, L.; Schlögl, S.; Groth, A. Linking Personality and Trust in Intelligent Virtual Assistants. *Multimodal Technol. Interact.* **2023**, *7*, 54. [CrossRef]
55. Buck, C.; Doctor, E.; Hennrich, J.; Jöhnk, J.; Eymann, T. General Practitioners’ Attitudes toward Artificial Intelligence-Enabled Systems: Interview Study. *J. Med. Internet Res.* **2022**, *24*, e28916. [CrossRef]
56. Chen, X.; Zou, D.; Xie, H.; Cheng, G.; Liu, C. Two Decades of Artificial Intelligence in Education. *Educ. Technol. Soc.* **2022**, *25*, 28–47.
57. Mäki, M.; Kauttonen, J.; Alamäki, A. How Students’ Information Sensitivity, Privacy Trade-Offs, and Stages of Customer Journey Affect Consent to Utilize Personal Data. 2023. Available online: <https://www.theseus.fi/handle/10024/793210> (accessed on 11 December 2024).
58. Nguyen, A.; Ngo, H.N.; Hong, Y.; Dang, B.; Nguyen, B.-P.T. Ethical Principles for Artificial Intelligence in Education. *Educ. Inf. Technol. Dordr* **2023**, *28*, 4221–4241. [CrossRef] [PubMed]
59. Roy, R.; Babakerkhell, M.D.; Mukherjee, S.; Pal, D.; Funilkul, S. Evaluating the Intention for the Adoption of Artificial Intelligence-Based Robots in the University to Educate the Students. *IEEE Access* **2022**, *10*, 125666–125678. [CrossRef]
60. Bilquise, G.; Ibrahim, S.; Salhieh, S.M. Investigating Student Acceptance of an Academic Advising Chatbot in Higher Education Institutions. *Educ. Inf. Technol. Dordr* **2023**, *29*, 6357–6382. [CrossRef]
61. Alamäki, A. A Conceptual Model for Knowledge Dimensions and Processes in Design and Technology Projects. *Int. J. Technol. Des. Educ.* **2018**, *28*, 667–683. [CrossRef]
62. Garrison, D.R.; Kanuka, H. Blended Learning: Uncovering Its Transformative Potential in Higher Education. *Internet High. Educ.* **2004**, *7*, 95–105. [CrossRef]

63. Zilka, G.C.; Cohen, R.; Rahimi, I. Teacher Presence and Social Presence in Virtual and Blended Courses. *J. Inf. Technol. Educ. Res.* **2018**, *17*, 103.
64. Lan, Y.-J. Immersion, Interaction, and Experience-Oriented Learning: Bringing Virtual Reality into FL Learning. *Lang. Learn. Technol.* **2020**, *24*, 1–15.
65. Heath, R.; Brandt, D.; Nairn, A. Brand Relationships: Strengthened by Emotion, Weakened by Attention. *J. Advert. Res.* **2006**, *46*, 410–419. [[CrossRef](#)]
66. Song, Y.; Dai, X.-Y.; Wang, J. Not All Emotions Are Created Equal: Expressive Behavior of the Networked Public on China's Social Media Site. *Comput. Hum. Behav.* **2016**, *60*, 525–533. [[CrossRef](#)]
67. Berger, J.; Milkman, K.L. What Makes Online Content Viral? *J. Mark. Res.* **2012**, *49*, 192–205. [[CrossRef](#)]
68. Lazarus, R.S. Cognition and Motivation in Emotion. *Am. Psychol.* **1991**, *46*, 352. [[CrossRef](#)] [[PubMed](#)]
69. Sheeran, P.; Webb, T.L. The Intention–Behavior Gap. *Soc. Pers. Psychol. Compass* **2016**, *10*, 503–518. [[CrossRef](#)]
70. Alamäki, A.; Pesonen, J.; Dirin, A. Triggering Effects of Mobile Video Marketing in Nature Tourism: Media Richness Perspective. *Inf. Process. Manag.* **2019**, *56*, 756–770. [[CrossRef](#)]
71. Kavanagh, S.; Luxton-Reilly, A.; Wuensche, B.; Plimmer, B. A Systematic Review of Virtual Reality in Education. *Themes Sci. Technol. Educ.* **2017**, *10*, 85–119.
72. Mikropoulos, T.A.; Natsis, A. Educational Virtual Environments: A Ten-Year Review of Empirical Research (1999–2009). *Comput. Educ.* **2011**, *56*, 769–780. [[CrossRef](#)]
73. Hall, E.; Seyam, M.; Dunlap, D. Identifying Usability Challenges in AI-Based Essay Grading Tools. In Proceedings of the International Conference on Artificial Intelligence in Education, Tokyo, Japan, 3–7 July 2023; pp. 675–680.
74. Valkokari, K.; Rantala, T.; Alamäki, A.; Palomäki, K. Business Impacts of Technology Disruption—a Design Science Approach to Cognitive Systems' Adoption within Collaborative Networks. In Proceedings of the Collaborative Networks of Cognitive Systems: 19th IFIP WG 5.5 Working Conference on Virtual Enterprises, PRO-VE 2018, Proceedings 19, Cardiff, UK, 17–19 September 2018; Springer International Publishing: Cham, Switzerland, 2018; pp. 337–348.
75. Sung, E.; Mayer, R.E. Affective Impact of Navigational and Signaling Aids to E-Learning. *Comput. Hum. Behav.* **2012**, *28*, 473–483. [[CrossRef](#)]
76. Cheung, M.F.Y.; To, W.-M. The Influence of the Propensity to Trust on Mobile Users' Attitudes toward in-App Advertisements: An Extension of the Theory of Planned Behavior. *Comput. Hum. Behav.* **2017**, *76*, 102–111. [[CrossRef](#)]
77. Shroff, R.H.; Deneen, C.C.; Ng, E.M.W. Analysis of the Technology Acceptance Model in Examining Students' Behavioural Intention to Use an e-Portfolio System. *Australas. J. Educ. Technol.* **2011**, *27*. [[CrossRef](#)]
78. Pavlou, P.A.; Gefen, D. Building Effective Online Marketplaces with Institution-Based Trust. *Inf. Syst. Res.* **2004**, *15*, 37–59. [[CrossRef](#)]
79. Rosseel, Y. lavaan: An R Package for Structural Equation Modeling. *J. Stat. Softw.* **2012**, *48*, 1–36. Available online: <http://www.jstatsoft.org/v48/i02/> (accessed on 11 December 2024). [[CrossRef](#)]
80. MacCallum, R.C.; Widaman, K.F.; Zhang, S.; Hong, S. Sample Size in Factor Analysis. *Psychol. Methods* **1999**, *4*, 84. [[CrossRef](#)]
81. Worthington, R.L.; Whittaker, T.A. Scale Development Research: A Content Analysis and Recommendations for Best Practices. *Couns. Psychol.* **2006**, *34*, 806–838. [[CrossRef](#)]
82. Mundfrom, D.J.; Shaw, D.G.; Ke, T.L. Minimum Sample Size Recommendations for Conducting Factor Analyses. *Int. J. Test.* **2005**, *5*, 159–168. [[CrossRef](#)]
83. Bürkner, P.-C.; Charpentier, E. Modelling Monotonic Effects of Ordinal Predictors in Bayesian Regression Models. *Br. J. Math. Stat. Psychol.* **2020**, *73*, 420–451. [[CrossRef](#)] [[PubMed](#)]
84. Silva, L.A.; Zanella, G. Robust Leave-One-out Cross-Validation for High-Dimensional Bayesian Models. *J. Am. Stat. Assoc.* **2024**, *119*, 2369–2381. [[CrossRef](#)]
85. Vehtari, A.; Gelman, A.; Gabry, J. Practical Bayesian Model Evaluation Using Leave-One-out Cross-Validation and WAIC. *Stat. Comput.* **2017**, *27*, 1413–1432. [[CrossRef](#)]
86. Bürkner, P.-C.; Vuorre, M. Ordinal Regression Models in Psychology: A Tutorial. *Adv. Methods Pract. Psychol. Sci.* **2019**, *2*, 77–101. [[CrossRef](#)]
87. Conijn, R.; Kahr, P.; Snijders, C. The Effects of Explanations in Automated Essay Scoring Systems on Student Trust and Motivation. *J. Learn. Anal.* **2023**, *10*, 37–53. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.