Understanding Trust in GenAI: The Roles of Reliability, Explanation Quality, and Confidence in Judgment

Fatima Bilal, Anssi Öörni, Jozsef Mezei
Åbo Akademi University, Turku, Finland
fatima.bilal@abo.fi, anssi.oorni@abo.fi, jozsef.mezei@abo.fi

Generative Artificial Intelligence (GenAI) is increasingly utilised in academic contexts; however, concerns about reliability and transparency raise questions about user trust and adoption. Drawing on Trust in Technology Theory, this study examines how perceived reliability and explanation quality influence trust in GenAI and whether trust mediates their effects on intention to use AI in an academic context. This research study also examines the role of confidence in judgment as a user-level factor, using survey-based data from 262 participants and hierarchical regression with a mediation analysis. The results show that perceived reliability and explanation quality significantly enhance trust in AI. However, trust partially mediates their relationship with intention to use GenAI. In contrast, confidence in judgment does not significantly predict trust or intention, highlighting a conceptual distinction between users' self-confidence and trust in AI systems. The findings highlight the importance of perceived reliability and explanation quality in fostering calibrated trust and responsible adoption of GenAI by students and teachers in educational settings.

DOI
https://doi.org/10.18690/um.fov.4.2026.25

ISBN
978-961-299-152-4

Keywords:
trust in AI, perceived reliability, explanation quality, confidence in judgment, adoption intention, generative ai (gen AI)

University of Maribor Press

1 Introduction

Generative Artificial Intelligence (GenAI) is defined as an advanced branch of Artificial Intelligence (AI) that includes creating and refining content, including text, music, and images, based on the input provided by the user or data that is accessible from the internet. GenAI utilises deep learning models for the identification of structures and patterns of large datasets and produces content and results by using large dataset patterns (Miranda et al., 2025). After its emergence in the year 2018, GenAI has evolved and been adopted in academic settings and professional contexts as systems based on large language models are now used for decision support, generating ideas, summarising input, drafting, and producing final results. The new and improved algorithms of AI, including reinforcement learning and deep learning, are able to process data in natural form and are capable of mining raw and unstructured data, including images and text (Dwivedi et al., 2023). Considering the academic settings, the use of GenAI has been widely adopted and has transformed academic learning; however, GenAI offers both opportunities and challenges in the development of students, impacting the level of users' trust in GenAI (Vieriu et al., 2025).

Trust is characterised as the key factor in human-AI interaction, as it refers to users' expectations considering how the AI system will generate results and outputs under risks on uncertain conditions (Vereschak et al., 2021). Past researchers (Lee & See, 2004; Lockey et al., 2021) have identified and discussed key factors of users' trust in technology by reviewing empirical methodologies, including reliability and explainability of the system. In an educational context and setting, the factor of AI explanation quality is important as students rely on AI-based output and results instead of its critical evaluation. The current literature on the use of GenAI has emphasised that students need to have a clear understanding and capability for successful use of GenAI in learning (Bećirović et al., 2025). However, the existing research on GenAI only lacks integration of user-level and system-level factors and examines trust cues solely. Similarly, the current research, including (Ursavaş, 2025; Zhou, 2024), is based on limited testing and examination of explanation quality, perceived reliability, and users' trust and confidence in GenAI, which also creates doubt in comprehension and understanding of the role of users' confidence. Existing empirical studies on trust in GenAI, including Rittenberg & Holland (2024), suggested that a higher level of self-confidence of AI users might result in calibrated

trust in automated systems. For addressing this gap, this research study aims to investigate how perceived reliability and explanation quality impact trust in GenAI and how trust mediates the impact of perceived reliability and explanation quality in terms of intention to use GenAI. This research also examines how confidence in AI judgments predicts the usage intention and trust of users.

2 Theoretical Background

2.1 Trust in Generative AI in Academic Contexts

As stated by the study of Lim et al. (2023), GenAI generates humanised content in response to the input of prompts and commands by the user by using large-scale deep learning models. In the academic context, the use of GenAI tools is significantly increasing for writing assessments, carrying out research synthesis, and retrieving information. However, the output produced by GenAI is subject to inaccuracies and errors, which result in concerns related to the responsible use of AI and reliability. In line with this, the research by Lee & See (2004) stated that trust plays a key role in examining how users can rely on AI systems and other factors that impact the use of AI systems by the users. In support of this notion, the research by Siau & Wang (2018) reviewed and examined trust in machine learning, robotics, and AI. The study reflected that trust is characterised as a key factor in determining the acceptance of AI, along with its development and continuous development in AI. In academic contexts, over-trust in AI links to acceptance of inaccurate outputs; however, under-trust might result in preventing users from receiving AI assistance. The research by Gilkson & Wooley (2020) reflected on the trust of human users in the use of technology and stated that initial research on the acceptance and use of new digital tools and technology relied more on user reactions to new technological features. Therefore, critical understanding and implications of trust in GenAI explain the overall academic level adoption and behaviour of users in GenAI in the academic context.

2.2 System-Based Antecedents of Trust

The study by McKnight et al. (2002) examined the integrative model proposed by McKnight, which incorporated trust types including disposition to trust, trusting beliefs, and trusting intentions. This framework was related to the framework of the

Theory of Reasoned Action (TRA), which resulted in beliefs that form attitude, which creates behavioural intentions and forms overall behaviour. The overall trust in TRA in GenAI is related to the reliability and behaviour of the system. In AI-based systems, competence is measured in terms of how the system is able to perform reliably and generate output and results in accurate terms, which are the system-based antecedents of the trust in the AI system by the users.

2.2.1 Perceived Reliability

Perceived reliability of an AI system is defined as the construct that impacts the level of trust of AI users, as it evaluates the consistency and accuracy of the output and results generated by AI systems. Perceived reliability in research, including Lee & See (2004), is related to two key factors: predictability and competence of the system. Therefore, in terms of GenAI research, users are more likely to perceive GenAI as reliable when predictability and competence level are higher, which results in users' willingness to rely on the system. In line with this, the recent study by Nelson et al. (2025) suggested that students are using AI after assessing key risks and benefits associated with the use of GenAI in academic writing. The study reflected that besides factors including perceived reliability and competence of GenAI tools, including ChatGPT, they are under constant pressure to achieve high grades, which makes them attempt to use GenAI in academic tasks and assessments. Therefore, the accuracy of GenAI is the key determinant in shaping the perceived reliability of the users.

2.2.2 Explanation Quality

In addition to perceived reliability, explanation quality is characterised as another key factor in shaping users' perceptions of AI systems. Explanation quality in GenAI refers to the comprehension and clarity in the overall system reasoning to generate output and results. Therefore, a high level of explanation results in increasing overall transparency, the system's reasoning and comprehension for AI users, and reducing ambiguity in data and results. The study by Hoff and Bashir (2015) reviewed factors that influence trust and stated that when AI users are able to comprehend and understand how the AI system works and how results and outputs are generated, they are more likely to refer to AI as credible and trustworthy. In support of this, the recent research by Hadra et al. (2026) reviewed that accuracy and reliability are

significant in terms of the use of GenAI in the academic context, as students are more likely to use results and output generated by GenAI when the output is non-plagiarized and authentic.

2.3 User-Based Antecedents: Confidence in Judgment

Besides system characteristics and trust of the user in GenAI, user-based antecedents are also critical in influencing trust in AI by confidence in the overall judgement regarding the produced output and results. Confidence in judgment by an AI user is the belief of an individual in their ability to evaluate whether the output generated by AI is accurate and reliable or not. In alignment with previous research, including Hoff & Bashir (2015) and Lee & See (2004), trust is defined as the willingness of users to rely on technology and systems, whereas confidence level is related to self-assurance and evaluation capability of the user in terms of AI-generated output. In line with this, Li et al. (2025) stated that there exists a complex relationship between confidence level and trust of the user in AI, as accuracy and verification of AI output and results lead to a higher level of trust; however, a high level of self-confidence results in reduced reliance, critical evaluation, and overall reliability. Therefore, examining the relationship between trust in AI and judgment requires user-level understanding in the context of system-level cues that shape overall trust.

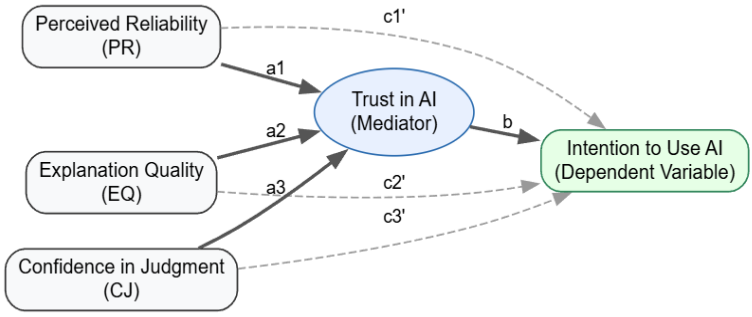

Figure 1: Conceptual Model

3 Hypotheses

As reported by Nazaretsky et al. (2025), trust performs a key role in determining whether users can rely on and adopt AI output and results in academic contexts. Prior research in the context of human and AI interaction stated that trust acts as a mechanism by which system characteristics influence the behavioural intention of AI users.

3.1 Perceived Reliability

Perceived reliability reflects users' belief that the AI system produces accurate and consistent outputs. Reliability signals competence and predictability, which are foundational elements of trust in automation. When users perceive GenAI as reliable, uncertainty is reduced, and trust is strengthened.

H1: Perceived reliability is positively associated with trust in AI.

H2: Trust in AI mediates the relationship between perceived reliability and intention to use AI in academic work.

3.2 Explanation Quality

Because generative AI systems often operate as "black boxes," explanation quality is a key trust-building cue. Clear and understandable explanations enhance transparency and reduce ambiguity, thereby strengthening users' trust in the system.

H3: *Explanation quality is positively associated with trust in AI.*

H4: *Trust in AI mediates the relationship between explanation quality and intention to use AI in academic work.*

Confidence in judgment refers to users' belief in their ability to evaluate AI outputs. The relationship between confidence and trust is theoretically ambiguous: confident users may either trust AI more (due to perceived evaluative ability) or less (due to increased scrutiny). Therefore, confidence is examined as an exploratory predictor of trust.

H5 (Exploratory): Confidence in judgment is associated with trust in AI. (See Appendix: 1).

4 Methodology

4.1 Research Design

This research study utilised a scenario-based online survey design for examining how users evaluate GenAI in academic contexts. The study did not measure the general attitude of AI users towards AI, but it arranged a standardised AI-based interaction for the participants before responding to survey questions. The arrangement of AI-based interaction was carried out to ensure that all research participants evaluated the AI output under controlled conditions, which led to strengthening internal validity. Under controlled conditions, research participants were provided with academic prompts, including a writing task and research-related claims accompanied by a GenAI response, which comprised a short and brief explanation of the integration and reasoning of GenAI in the academic context. Research participants were asked to read the provided prompts and AI response in detail by supposing that they had used the system for real and practical academic purposes. After this, the research participants utilised their own judgement in terms of evaluating AI reliability, explanation quality, and trust in AI regarding academic tasks. This post-task evaluation approach follows established practices in human–AI interaction research, where perceptions and trust are measured immediately after system exposure.

Although the study adopts a cross-sectional survey design, the inclusion of a standardized scenario introduces a quasi-experimental element by ensuring that all participants evaluate the same AI-generated output under controlled conditions. This approach allows for consistency in stimulus exposure while capturing participants' perceptions in a realistic decision-making context. A fully manipulated experimental design was not implemented in this study in order to prioritise ecological validity and to capture naturally occurring variations in user perceptions. Future research could extend this approach by experimentally manipulating system characteristics such as reliability levels or explanation styles to establish stronger causal inference.

4.2 Participants and Sampling

The total number of research participants recruited under this study was 262 (*N=262*), and they were recruited by use of convenience sampling from social media channels and university study groups for understanding their trust in GenAI. The total sample of 262 participants comprised young adults and university students, and they were from different fields, including social science, engineering, and business, and mainly from three countries: Sweden, the United Kingdom (UK), and Finland. Research participants were kept well-informed regarding their voluntary participation and their anonymity throughout the research process and data analysis. Before the collection of data, informed consent was obtained from all participants in electronic form. In the survey questionnaire, the basic demographic information collected from participants included their gender, field of study/education, and their current highest degree level. A seven-point scale was arranged in the survey questionnaire for respondents to indicate their previous experience of use of AI technologies, including 1=very low experience and 7=very high experience. The research constructs and variables were added in a subsequent section of the research survey in this study.

4.3 Measures

In the research study, all the specified constructs were measured by use of a seven-point Likert scale in which a low score indicated limited endorsement and relevance of the construct, whereas a high score indicated stronger endorsement of the constructs in the research study. All measurement items were adapted from prior validated scales in the trust in technology and AI literature (e.g., McKnight et al., 2002; Hoff & Bashir, 2015; Lee & See, 2004), with minor wording adjustments to fit the GenAI academic context.

4.3.1 Trust in AI

Trust in AI was measured in the research study by use of a multi-item scale which was adopted under the GenAI academic context (For example. *"I believe that GenAI provides reliable output, and I completely trusted using the AI-provided recommendations".* The level of Cronbach's $\alpha = .70$, indicating acceptable internal consistency.

4.3.2 Intention to Use AI

Intention to use AI was measured in this research study as a multi-item scale, which was adopted from existing research in terms of technology acceptance and adoption (For example, *"I would use GenAI in real academic tasks"*. The level of Cronbach's α was .70 which indicates that internal consistency in terms of the intention of using AI was acceptable.

4.3.3 Perceived Reliability

Perceived reliability was measured in this research study by the use of a single-item indicator, which utilised the perception of participants in terms of accuracy and consistency of AI (For example, "*I perceived that AI was reliable in providing results*") The rationale for utilising perceived reliability as single-item measures which were allocated in controlled conditions when construct variable was observed in the specified task.

4.3.4 Explanation Quality

Explanation quality in this research for examining trust in the use of GenAI was examined by the use of a single item by evaluation of the AI reasoning in terms of its usability and clarity (For example, *"GenAI was able to explain the reasoning clearly")*. This construct of explanation quality was able to examine the evaluation of participants in terms of the provided explanation in the standardised response of GenAI.

4.3.5 Confidence in Judgment

Confidence in judgment was measured using a single-item self-assessment (For example, "*I am confident in my ability to evaluate whether the AI's answer is correct*"), which was rated from 1 (not at all confident) to 7 (very confident).

4.4 Data Analysis

In this research study, data analysis was performed through the use of statistical software IBM SPSS (Statistical Package for Social Sciences) Version 28. The data collected from 262 research respondents were included in the dataset and analysed

using SPSS. Initially, descriptive statistics and correlation tests were done for the evaluation of initial relationships between the research variables. Secondly, multiple regression analysis was performed hierarchically for the assessment of intention to use GenAI by predictors and constructs. In the provided Model 1, demographic controls included AI experience of users, gender, current degree level, and field of study. In Model 2, the explanatory power of research variables was examined in terms of constructs including perceived reliability, explanation quality, and trust in AI. For examining the mediating role of trust in the developed hypothesis, the researcher carried out mediation analysis using Hayes's PROCESS macro (Model 4) by inclusion of 95% confidence intervals and 5,000 bootstrap samples. The construct of trust in AI was characterised as a mediator between the perceived reliability of GenAI and intention to use, and explanation quality and intention to use AI by the user. In the research hypothesis testing, the mediation model included control variables, which were statistically analysed by use of a two-tailed test, and the significance level was evaluated at an estimation level of 0.05. Research participants' evaluation of AI output in controlled conditions was used in single-item indicators for assessing perceived reliability and explanation quality. The psychological constructs, including trust and user intention to use GenAI, were measured in this research by the development of multi-item scales. For minimising the risk of bias, participants examined the provided AI scenario before responding to the survey questions. The use of Harman's Single-factor test in the analysis indicated that the majority of the variance (<40%) did not result from a single factor, which resulted in the common method variance being limited in terms of explaining the relationship between all the included variables in the research study.

5 Empirical Findings

5.1 Descriptive Statistics

The first section of empirical findings discusses descriptive statistics, which were analysed under this research study. Table 2 in the provided appendix section summarises that research respondents' responses with a low to moderate level of trust in AI, including a mean (M) = 2.55 and a standard deviation (SD) value of 1.49. The level of perceived reliability resulted in M=2.70 and SD=1.72. Similarly, the observed intention to use AI by users in academic work was low, with M=2.43 and SD=1.79. For the following construct of explanation quality, the results were

M=2.43 and SD=1.90, which suggested that explanation quality is a critical factor and GenAI can be used in academic tasks after caution and diligence (See Appendix 1).

5.2 Bivariate Correlations

Under bivariate correlation analysis, Pearson Correlation was performed, which indicated that strong user intention to use AI in academic contexts was one of the key factors in the overall positive perception of users associated with the use of GenAI. Intention to use GenAI was positively related to trust in AI ($r=.40$, $p<0.001$), explanation quality ($r=.64$, $p<0.001$), and perceived reliability ($r=.60$, $p<0.001$). Results suggested that the reliability of GenAI and explanation quality are related to forming trust. These results are consistent with past research on AI experience with calibration perspective having intention to use AI ($r=-45$, $p<0.001$), which indicated that the higher the level of user experience in GenAI, the more critical evaluation of the AI system is carried out (See Appendix: 3).

5.3 Multiple Regression

A hierarchical multiple regression analysis was performed to assess the predictive capacity of demographic and AI perception variables on the utilisation of AI in the academic context. Model 1, comprising solely demographic variables (gender, field, degree, experience with AI), accounted for 28% of the variance in the intention to use AI, $R^2 = .280$, $F(4, 256) = 24.91$, $p < .001$. Key predictors were field of study ($\beta = .154$, $p = .004$), degree level ($\beta = .240$, $p < .001$), and familiarity with AI ($\beta = -.379$, $p < .001$) (See Appendix: 4). Model 2, incorporating variables such as trust, reliability, and explanation quality, accounted for an additional 28.5% of the variance, resulting in a total explained variance of 56.5%, $R^2 = .565$, $F(8, 252) = 40.96$, $p < .001$. Among the new predictors, reliability ($\beta = .238$, $p < .001$), explanation quality ($\beta = .233$, $p < .001$),. This suggests that perceptions of perceived reliability and explanation quality are central determiners of academic use of AI tools, surpassing the influence of demographic factors (See Appendix: 5)

Although trust in AI was included in a final model, the impact was not significant ($\beta = .062$, $p = .218$), which shows that it does not uniquely predict the intention to use AI. The ANOVA confirmed that both models were statistically significant: Model

1, $F(4, 256) = 24.91$, $p < .001$; Model 2, $F(8, 252) = 40.96$, $p < .001$. This indicates that the regression models reliably predicted intention to use AI. Trust in AI did not remain a statistically significant independent predictor of intention in the regression model; this result likely reflects shared variance between trust, perceived reliability, and explanation quality. Variance inflation factors (VIFs) ranged between 1.4 and 2.3, indicating no critical multicollinearity concerns. Mediation analysis isolates indirect pathways rather than unique direct contributions, confirming that trust functions as a transmission mechanism linking system attributes to the behavioural intention of the AI user.

5.4 Mediation analyses (PROCESS Model 4, 5,000 bootstraps)

For examining how trust in AI mediates the relationship between confidence judgment and intention of using AI, we used 5,000 bootstrap samples and a 95% confidence interval to run Hayes' PROCESS macro (Model 4; Hayes, 2022). The independent variable (X) was confidence judgment; the mediator (M) was trust in AI; and the dependent variable (Y) was intention to use AI. The path from confidence judgment to trust (a-path) was not significant, nor was the direct effect on intention. The indirect effect (a×b) was also non-significant as the CI includes zero, indicating no mediation. Thus, did not influence intention via trust. Therefore, Hypothesis 5 is not supported (See Appendix: 6).

Trust positively mediates the association between perceived reliability and intention. The direct effect remains strong and significant, indicating partial mediation. Hypothesis 2 is supported (see Appendix 7). Significant partial mediation: Clearer reasoning builds Trust and directly boosts Intention; both routes matter. Perceived explanation clarity significantly increases trust, which in turn predicts intention. The indirect effect is large, while the direct effect remains robust, indicating partial mediation, which results in hypothesis 3 being supported (See Appendix: 8).

6 Discussion

The findings of this study should be interpreted as associative rather than strictly causal due to the cross-sectional nature of the research design. The results obtained from data analysis indicated that perceived reliability and explanation quality are the key variables in determining user trust in GenAI in academic contexts. The findings

obtained under this research study are consistent with the past research (Lee & See, 2004; McKnight et al., 2002), which also related that when accurate and consistent outputs and results are generated by GenAI, users are more likely to develop trust level and use AI in academic contexts. Therefore, reliability and transparency of the AI system are characterised as key antecedents of trust in the use of GenAI. Similarly, the findings also resulted in a mean level of trust and intention to use AI, which suggested that users are more cautious in using GenAI in academic settings due to concerns related to the reliability of results and explanation quality. Considering educational and academic contexts, the responsible use and adoption of AI is related to the transparency and ability of the AI system to generate results.

The findings obtained under this research study indicate that trust is a key element in the academic context with relation to the use of GenAI, and AI users are responsible for carrying out final decisions after generating results and output from GenAI. In an academic context, explanation quality reduces the level of uncertainty and increases the level of transparency for the AI users. These findings align with the research by Siau & Wang (2018), which also stated that reliability and explanation quality result in forming behavioural intention of the AI users. Similarly, under the data analysis, when explanation quality and reliability were tested simultaneously in the regression model, trust resulted as a transmission mechanism, which is accompanied by adoption intention to use GenAI output and results. The mediation analysis performed in this research study resulted in the finding that trust is positively associated with both explanation quality and perceived reliability for the user's intention to use AI in academic contexts. Therefore, the findings obtained from this research support that trust is characterised as a mechanism that forms a relationship between the AI system and the intended user behaviour.

The findings from this research indicate that, in academic contexts, reliability and explanation quality are likely to enhance AI users' intention to use AI systems and to build trust. These findings align with the research of Li & Aral (2025), which examined how the level of trust affects overall engagement among AI users. Users are more likely to use GenAI output and results if they find it reliable and usable academically. Similarly, for the other construct, including confidence in judgement, it has been found that confidence in judgement by the AI user does not directly result in the prediction of intention to use AI or overall trust level, which suggests the ability of AI users in their own evaluation does not always result in generating a

higher level of trust in provided AI systems and tools. This theoretical and practical distinction is critical as trust relates to the willingness of users to rely on an external system, whereas confidence is related to the user and his/her own ability to use AI. These findings supported the past research carried out by Agudo et al. (2024) that even when users who are fully aware and confident in their own analysis and judgment are not always likely to rely on AI systems and their results. Therefore, the self-efficacy of AI users is different from system trust, and user-level confidence is increased, and trust is formed in academic contexts when system-level cues, including explanation quality and perceived reliability, are higher.

The findings of this research study also emphasise the significance of calibrated trust of AI users and how the trust level is increased and promoted by enhancing perceived reliability and explanation quality. The results obtained reflected that developers and analysts are well aware of the fact that improvement in perceived reliability and explanation quality does not always result in increased reliance and use of the AI system due to self-efficacy and self-confidence of the users. This is in alignment with the research carried out by Ding et al. (2023) that AI users may develop over-reliance on AI systems in the absence of self-efficacy. Therefore, AI systems are likely to be trusted and utilised when the trust level is positively impacted by perceived reliability and explanation quality.

7 Summary, Implications, and Outlook

The findings demonstrated that perceived reliability and explanation quality are central determinants of trust. When AI systems provide consistent outputs and clear reasoning, users are more likely to develop trust and express an intention to integrate these tools into academic tasks. In contrast, confidence in judgment did not significantly influence trust in AI. This finding highlights an important theoretical distinction: trust in AI and confidence in one's own evaluative ability are related but conceptually distinct constructs. Trust reflects a willingness to rely on the system, whereas confidence reflects belief in one's own competence. The absence of a significant relationship suggests that enhancing system transparency and reliability may be more influential for adoption than strengthening users' self-confidence alone.

With the significant rise in use of GenAI in the academic context, the ability of users to produce results and output from AI systems has significantly increased; however, it has been characterised as a key challenge for users to evaluate its quality and perceived reliability. This study aimed to examine the system-level constructs, including perceived reliability of GenAI and explanation quality, which increases trust in AI and also increases the intention of the user to utilise GenAI in academic work. The findings obtained under this research study revealed that perceived reliability and explanation quality of the AI system determine the level of trust of users in GenAI consistency in AI-based output results in the development of trust and users' intention to use GenAI. However, confidence of the user in their own judgement and evaluation does not impact the level of trust in AI. Therefore, for increasing the reliance on GenAI output and its usability, it is important to enhance the overall system transparency and reliability to generate accurate results.

7.1 Theoretical Contributions

In terms of theoretical contribution, this research study contributed to the GenAI context in three key ways. Firstly, it provided a clear difference between the perceived reliability (competence) and explanation quality (transparency) system cues and how they form trust level in an academic context in performing tasks and assessments. Secondly, the findings of this research study also emphasised comprehension and implications of trust as a mechanism for AI users and how it mediates the influence of perceived reliability and explanation quality. Thirdly, this research study was able to draw a conclusion that trust in AI is conceptually distinct from confidence in personal evaluation and judgment, and self-confidence is not directly associated with the usability and trust prediction in the academic context.

7.2 Limitations

Irrespective of its critical insights and findings on trust in the use of GenAI in the academic context, this research study has certain limitations. The sample size of 262 participants was obtained through convenience sampling and was restricted to students and academic users, which limited the generalizability of this research to other AI users. Secondly, the research study made use of self-reported measures of AI, including trust level, perceived reliability, and intention of the user to use AI. The self-reported measures are prone to common method variance and social bias,

which could have impacted the results of this research. Thirdly, the controlled standardised protocol and design might have impacted the long-term development of trust in the research study for the AI users. An additional limitation relates to potential omitted variables such as personality traits (e.g., openness, conscientiousness) or individual differences in risk perception, which may influence both trust in AI and intention to use. Similarly, cultural factors such as uncertainty avoidance or technological familiarity could shape user perceptions. Future studies should incorporate these variables to better isolate the effects observed in this research. Although the sample includes participants from multiple countries, it remains relatively homogeneous and limited in size. Future research should employ larger and more culturally diverse samples to enhance generalizability and capture global patterns of AI trust and adoption.

7.3 Future Research Directions

In terms of future research on trust in the use of GenAI in the academic context, researchers and practitioners are required to perform an in-depth examination of how AI forms trust in terms of users' cultural attitudes, educational systems, and institutional norms. Researchers can also perform comparative analysis-based studies to understand c differences in terms of trust in AI and trust calibration. Similarly, longitudinal research can be carried out to examine how reliance, evaluation, judgment, and level of trust are evolved over time and how AI users acquire experience and insights in GenAI systems. Future research studies are also required to investigate the allocation of AI functions and users' decisions in responsible terms in an AI-assisted environment. Thus, understanding of AI-based recommendations and judgment will help in the development of tools and models in optimising the level of trust calibration and reducing over- and under-reliance on GenAI in the academic context. Future research could also compare different generations of AI models (e.g., earlier vs. more advanced versions) to examine how improvements in model capability influence user trust and reliance.

Proofreading statement:

The authors confirm that this manuscript has been carefully proofread and revised for language quality, grammar, and overall clarity prior to submission.

References

Bećirović, S., Polz, E., & Tinkel, I. (2025). Exploring students' AI literacy and its effects on their AI output quality, self-efficacy, and academic performance. *Smart Learning Environments*, *12*(1), 29. https://link.springer.com/article/10.1186/s40561-025-00384-3

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., et al. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges, and implications of generative conversational AI for research, practice, and policy. International Journal of Information Management, 71, 102642

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Hadra, M., Cambridge, K., & Mesbah, M. (2026). Evaluating the accuracy and reliability of AI content detectors in academic contexts. *International Journal for Educational Integrity*, *22*(1), 4.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Li, H., & Aral, S. (2025). Human trust in AI search: A large-scale experiment. *arXiv preprint*. https://arxiv.org/abs/2504.06435

Li, J., Yang, Y., Liao, Q. V., Zhang, J., & Lee, Y. C. (2025, April). As confidence aligns: understanding the effect of AI confidence on human self-confidence in human-AI decision making. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). https://doi.org/10.1145/3706598.3713336

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. The International Journal of Management Education, 21(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research, 13*(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

Miranda, F. J., & Chamorro-Mera, A. (2025). Exploring the adoption of generative artificial intelligence tools among university teachers. *Higher Education Research & Development*, 1–17. https://doi.org/10.1080/07294360.2025.2559648

Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T. (2025). The critical role of trust in adopting AI-powered educational technology for learning: An instrument for measuring student perceptions. *Computers and Education: Artificial Intelligence*, *8*, 100368. https://doi.org/10.1016/j.caeai.2025.100368

Nelson, A. S., Santamaría, P. V., Javens, J. S., & Ricaurte, M. (2025). Students' perceptions of generative artificial intelligence (GenAI) use in academic writing in English as a foreign language. *Education Sciences*, *15*(5), 611.

Rittenberg, B. S., Holland, C. W., Barnhart, G. E., Gaudreau, S. M., & Neyedli, H. F. (2024). Trust with increasing and decreasing reliability. *Human Factors, 66*(12), 2569–2589. https://doi.org/10.1177/00187208231200471

Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *ACM SIGMIS Database, 49*(4), 99–104. https://doi.org/10.1145/3242734.3242753

Ursavaş, Ö. F., Yalçın, Y., İslamoğlu, H., Bakır-Yalçın, E., & Cukurova, M. (2025). Rethinking the importance of social norms in generative AI adoption: Investigating the acceptance and use of generative AI among higher education students. *International Journal of Educational Technology in Higher Education, 22*(1), 38. https://doi.org/10.1186/s41239-025-00467-3

Vereschak, O., Bailly, G., & Caramiaux, B. (2021). How to evaluate trust in AI-assisted decision-making? A survey of empirical methodologies. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–39. https://doi.org/10.1145/3476068

Vieriu, A. M., & Petrea, G. (2025). The impact of artificial intelligence (AI) on students' academic development. *Education Sciences*, *15*(3), 343. https://doi.org/10.3390/educsci15030343

Zhou, T., & Lu, H. (2024). The effect of trust on user adoption of AI-generated content. *Electron. Libr.*, 43, 61-76. https://doi.org/10.1108/el-08-2024-0244

Appendixes

Appendix: 1

Table 1. Hypothesis development.

Hypothesis	Statement
H1	Perceived reliability positively is associated with trust in AI.
H2	Trust in AI mediates the relationship between perceived reliability and intention to use AI in academic work.
H3	Explanation quality positively is associated with trust in AI.
H4	Trust in AI mediates the relationship between explanation quality and intention to use AI in academic work.
H5 (Exploratory)	Confidence in judgment is associated with trust in AI.

Appendix: 2

Table 2: Descriptive Statistics (N = 262)

Variable	M	SD
Trust in AI	2.55	1.49
Perceived Reliability	2.70	1.72
Explanation Quality	2.43	1.90
Intention to Use AI	2.43	1.79

Appendix: 3

Table 3: Correlations Among Study Variables

Variable	1	2	3	4	5	6	7
1. Gender	—						
2. Field	.078	—					
3. Experience with AI	.140*	−.083	—				
4. Trust in AI	.132*	.054	−.209**	—			
5. Perceived Reliability	−.046	.074	−.369**	.510**	—		
6. Explanation Quality	−.028	.201**	−.442**	.364**	.511**	—	
7. Intention to Use AI	−.063	.201**	−.450**	.404**	.596**	.637**	—

$p < .05$, ** $p < .01$

Appendix: 4

Table 4: Hierarchical Regression Summary

Predictor	Model 1 β (p)	Model 2 β (p)
Gender	.022 (.684)	—
Field of Study	.154 (.004)	—
Experience with AI	–.379 (<.001)	—
Degree Level	.240 (<.001)	—
Perceived Reliability	—	.238 (<.001)
Explanation Quality	—	.233 (<.001)
Trust in AI	—	ns†

Appendix: 5

Table 5: Anova Regression

Model Statistics	Model 1	Model 2
R^2	.280	.565
Adjusted R^2	.269	.551
Standard Error	1.531	1.200
F (df1, df2)	24.91 (4, 256)	40.96 (8, 252)
p-value	< .001	< .001

Appendix: 6

Table 6: Mediation analysis

Path	B (Unstd.)	SE	T	p	95% CI [LLCI, ULCI]
Path a: Confidence → Trust	–0.097	0.061	–1.59	.113	[–0.217, 0.023]
Path b: Trust → Intention	0.628	0.040	15.68	.000	[0.549, 0.707]
Direct Effect (c'): Confidence → Intention	–0.061	0.040	–1.55	.123	[–0.139, 0.017]
Indirect Effect (a×b)	–0.061	0.043	–	–	[–0.149, 0.020]

Appendix: 7

Table 7: Mediation analysis

Path	B	SE	T	P	95% CI
X → M	.4420	.0463	9.55	< .001	[.351, .533]
M → Y	.3479	.0308	11.29	< .001	[.287, .409]
Direct (c') X → Y	.4871	.0267	18.23	< .001	[.435, .540]
Indirect (a×b)	.1538	.0269	—	—	[.105, .209]

Appendix: 8

Table 8: (AI Explanation → Trust → IntentionExplanation →t partial mediation

Path	B	SE	t	P	95% CI
X → M	.2858	.0454	6.30	< .001	[.197, .375]
M → Y	.4274	.0245	17.41	< .001	[.379, .476]
Direct (c') X → Y	.4463	.0193	23.15	< .001	[.408, .484]
Indirect (a×b)	.1222	.0248	—	—	[.075, .173]

Table 9: Summary of Hypothesis

Hypothesis	Result
H1: Perceived reliability positively is associated with trust in AI.	Supported
H2: Trust mediates the relationship between perceived reliability and intention to use AI.	Supported (partial mediation)
H3: Explanation quality positively is associated with trust in AI.	Supported
H4: Trust mediates the relationship between explanation quality and intention to use AI.	Supported (partial mediation)
H5 (Exploratory): Confidence in judgment is associated with trust in AI.	Not supported