Determinants of AI Distrust Among Students in Innovation Learning

Prita Ayu Kusumawardhany[1,*] Natalia Cipto Mangun Kusumo[2]

[1] *Faculty of Business and Economics, Management Department, University of Surabaya, Indonesia*
[2] *Faculty of Business and Economics, International Business Networking, University of Surabaya, Indonesia*
[*]*Corresponding author. Email: pritaayu@staff.ubaya.ac.id*

ABSTRACT
This study investigates the effects of distrust in data protection, functional distrust, and distrust of AI replacing humans on students' distrust of Artificial Intelligence (AI) in an innovation learning context. This study employs a survey-based approach and partial least squares Structural Equation Modeling (PLS-SEM) with bootstrapping to examine the proposed hypothesis. A total of 104 responses from innovation students at the University of Surabaya were collected using a structured questionnaire. Findings reveal that all three hypotheses were supported. Functional distrust about performance skepticism emerged as the strongest predictor of AI distrust, followed by data privacy concern and human superiority belief, with the three variables collectively explaining a substantial proportion of the variance in AI distrust. Findings suggest that educators and AI developers should address students' capability skepticism and privacy concern to foster greater AI acceptance for innovation in higher education.
Keywords: AI Distrust, Data Protection, Functional Distrust, AI Replacing Humans, AI Innovation

1. INTRODUCTION

Artificial intelligence (AI) has emerged as one of the most transformative technologies of the Fourth Industrial Revolution, fundamentally reshaping how individuals learn, communicate, solve problems, and make decisions across numerous sectors. The rapid advancement of generative AI technologies, such as ChatGPT, Microsoft Copilot, Gemini, and Claude, has accelerated the integration of intelligent systems into educational environments (Chalmers et al., 2021). Universities worldwide increasingly adopt AI-powered learning platforms to enhance instructional quality, facilitate personalized learning, provide immediate feedback, generate educational content, and improve students' learning efficiency (Dwivedi et al., 2023). Consequently, AI is no longer viewed merely as a technological innovation but has become an essential component of modern higher education that supports innovation-driven learning and digital transformation (Smirnova et al., 2026).

Despite these remarkable advantages, the growing integration of AI into higher education has simultaneously introduced various psychological, ethical, and technological concerns. Rather than fully embracing AI technologies, many university students demonstrate hesitation, skepticism, and resistance toward relying on AI in academic environments. Recent studies indicate that AI adoption cannot solely be explained by positive perceptions such as perceived usefulness, perceived ease of use, or behavioral intention. Instead, students increasingly evaluate AI based on concerns regarding transparency, fairness, accountability, privacy protection, algorithmic reliability, and the possibility of replacing meaningful human interaction during learning. These concerns collectively contribute to the emergence of distrust, which represents an important psychological barrier to AI acceptance (Smirnova et al., 2026).

Unlike trust, which encourages individuals to engage with technology under conditions of uncertainty, distrust represents a distinct psychological construct characterized by suspicion, negative expectations, uncertainty, and protective behavior (Chen et al., 2024). Previous literature emphasizes that distrust should not simply be interpreted as the absence of trust because individuals may simultaneously recognize AI's potential benefits while remaining reluctant to depend on its recommendations (Vaassen, 2022). Distrust encourages users to become more cautious, critically evaluate AI-generated outputs, reduce system usage, or completely avoid interacting with AI technologies despite their availability. Therefore, understanding distrust provides deeper insight into technology resistance than traditional technology acceptance frameworks alone (Smirnova et al., 2026).

Supporting these educational findings, Kumari (2026) demonstrates that distrust also plays a critical role in determining user disengagement across AI-enabled service environments. In financial advisory services, users increasingly withdraw from AI systems because of perceived privacy risks, psychological discomfort, and feelings of creepiness associated with intelligent technologies. The study confirms that distrust functions as a mediating mechanism through which various technological and psychological risks ultimately reduce users' engagement with AI-enabled systems. These findings indicate that distrust is not limited to financial services but may represent a universal psychological mechanism influencing AI adoption across multiple application domains, including education.

Kumari (2026) further argues that technological innovations introducing uncertainty regarding privacy protection, algorithmic transparency, and ethical accountability are more likely to generate distrust than confidence. Users who perceive AI systems as opaque or difficult to understand become increasingly reluctant to rely on automated recommendations, even when those recommendations may improve performance or efficiency. The Theory of Distrust explains this phenomenon by proposing that insecurity, suspicion, and anticipated negative outcomes motivate individuals to avoid potentially risky technologies. Therefore, distrust serves as a critical explanatory mechanism for understanding why users disengage from AI despite recognizing its potential benefits (Tiwari et al., 2024).

2. RESEARCH METHOD

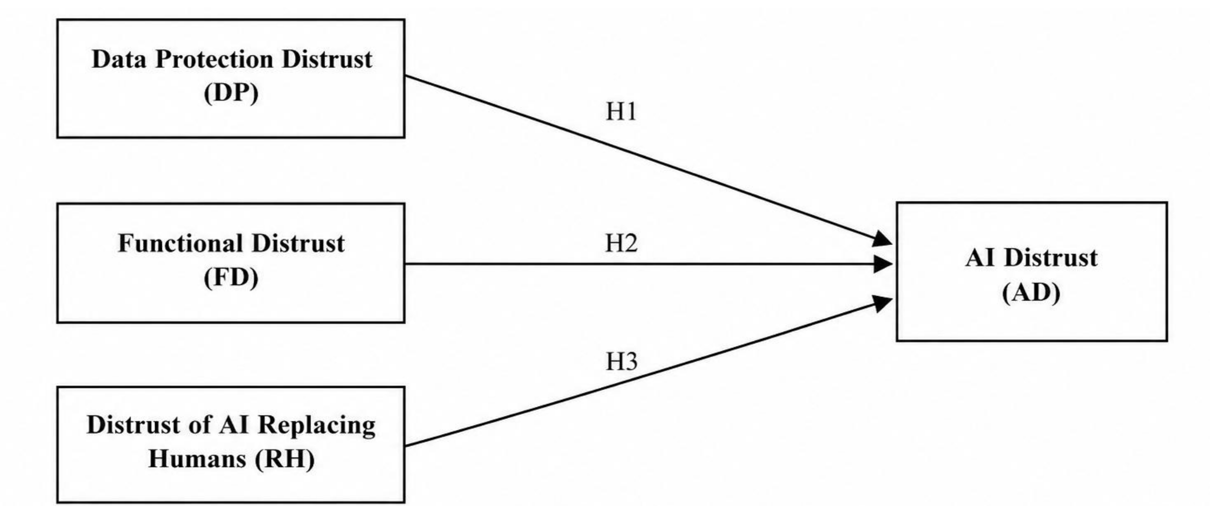
2.1 Measurement Scale

The measurement scale used in this study was adapted from previously validated studies on artificial intelligence adoption and AI distrust. All constructs were operationalized using multiple reflective indicators measured on a five-point Likert scale, ranging from 1 = Strongly Disagree to 5 = Strongly Agree. The use of a five-point Likert scale is considered appropriate because it enables respondents to express varying degrees of agreement while maintaining simplicity and reliability in survey responses.

The research model consists of four latent constructs: Data Protection Distrust (DP), Functional Distrust (FD), Distrust of AI Replacing Humans (RH), and AI Distrust (AD). Data Protection Distrust measures students' concerns regarding the privacy, security, and protection of personal data when using AI applications. Functional Distrust reflects students' perceptions of the reliability, accuracy, and overall performance of AI systems in supporting innovation learning. Distrust of AI Replacing Humans assesses students' concerns that artificial intelligence may substitute the essential roles of lecturers or trainers in providing guidance, feedback, ethical judgment, and practical learning experiences. AI Distrust represents students' overall skepticism and reluctance to fully rely on AI technologies in innovation-oriented learning activities (Smirnova et el., 2026; Kumari, 2026).

Each construct was measured using three indicators adapted from previous literature and modified to fit the context of university students enrolled in Innovation Management courses. Before conducting the structural model analysis, the measurement model was evaluated through reliability and validity assessments, including Cronbach's Alpha, Composite Reliability (CR), Average Variance Extracted (AVE), outer loadings, and the Heterotrait–Monotrait Ratio (HTMT). These procedures ensure that all measurement items adequately represent their respective constructs and satisfy the recommended criteria for PLS-SEM analysis (Hair et al., 2022; Sarstedt et al., 2024).

2.2 Research Model



3. RESULT AND DISCUSSION

This study examines the determinants of AI distrust among university students enrolled in Innovation Management courses using Partial Least Squares Structural Equation Modeling (PLS-SEM). The research specifically investigates how Data Protection Distrust (DP), Functional Distrust (FD), and Distrust of AI Replacing Humans (RH) influence students' overall AI Distrust (AD) within the context of innovation learning.

The analysis was conducted in two sequential stages. First, the measurement model was evaluated by assessing internal consistency reliability, convergent validity, and discriminant validity. The results indicate that all constructs achieved satisfactory Cronbach's Alpha, Composite Reliability, Average Variance Extracted (AVE), outer loading, and HTMT values, demonstrating that the measurement instrument is both reliable and valid for measuring students' perceptions of AI distrust.

After confirming the adequacy of the measurement model, the structural model was evaluated to examine the proposed hypotheses. The findings reveal that Data Protection Distrust, Functional Distrust, and Distrust of AI Replacing Humans collectively explain 47.7% of the variance in AI Distrust (R^2 = 0.477), indicating a moderate level of explanatory power. Furthermore, the model achieved a Q^2 value of 0.523, confirming that it possesses satisfactory predictive relevance in explaining students' distrust toward AI in innovation learning.

Among the four constructs included in the research model, Data Protection Distrust (DP) demonstrated the strongest indicator performance. Specifically, indicators DP1 and DP2 obtained the highest loading values (0.934), while DP3 also exhibited a very high loading value (0.931). These results indicate that students consistently perceive concerns regarding the protection, storage, and security of personal information as important determinants of distrust toward artificial intelligence. The high loading values further suggest that issues related to data privacy constitute one of the most consistently perceived dimensions of AI distrust among university students.

Overall, the findings suggest that students appreciate the benefits of artificial intelligence as a learning support tool, particularly for generating ideas, accessing information quickly, and assisting with innovation-related tasks. However, they remain reluctant to rely entirely on AI due to concerns about data privacy and security, the accuracy and reliability of AI-generated information, and the possibility that AI may replace the essential role of human educators. Among students taking Innovation Management courses, innovation learning is perceived as requiring not only technological assistance but also direct interaction with lecturers or trainers who can provide contextual explanations, practical insights, critical discussion, personalized feedback, and mentorship. Therefore, the results indicate that students prefer a collaborative learning approach in which AI serves as a complementary tool rather than a complete substitute for human instructors, highlighting the importance of maintaining human guidance in innovation-oriented education.

3.1 Hypothesis Testing

After confirming the explanatory power and predictive relevance of the structural model, the proposed hypotheses were evaluated using the bootstrapping procedure in SmartPLS 4. Bootstrapping was employed to estimate the statistical significance of the structural relationships among the latent constructs by calculating the standardized path coefficients (β), t-statistics, and p-values (Panahi et al., 2023). Following the recommendations of Hair et al. (2022), a hypothesis was considered statistically supported when the t-statistic exceeded 1.96 and the corresponding p-value was below 0.05.

The results of the hypothesis testing are summarized in Table 1. The table presents the standardized path coefficients, t-statistics, p-values, and the final decision for each proposed hypothesis.

Hipotesis	**β (Path Coeff)**	**Std. Error**	**t-statistik**	**p-value**	
H1: Data Protection Distrust → AI Distrust	0,323	0,096	**3,373**	**0,001**	**Accepted**
H2: Functional DIstrust → AI Distrust	0,666	0,063	**10,618**	**0**	**Accepted**
H3: Distrust of AI Replacing Humans → AI Distrust	0,321	0,098	**3,277**	**0,001**	**Accepted**
R^2 Model (AI Distrust) = 0.477 | Q^2 = 0.523 | sig. α = 0.05, t-value = 1.96					

Table 1. Hypothesis Testing Results

The results presented in Table 1 provide empirical evidence regarding the relationships between Data Protection Distrust, Functional Distrust, Distrust of AI Replacing Humans, and AI Distrust. The statistical significance of each relationship was evaluated based on the standardized path coefficient, t-statistic, and p-value. The interpretation of each hypothesis is discussed in the following subsections.

4. CONCLUSION

This study investigated the determinants of artificial intelligence (AI) distrust among university students in the context of innovation learning by examining the effects of Data Protection Distrust, Functional Distrust, and Distrust of AI Replacing Humans on overall AI Distrust. Using Partial Least Squares Structural Equation Modeling (PLS-SEM), the findings demonstrate that all three determinants significantly influence students' distrust toward AI technologies. These results indicate that AI distrust is a multidimensional phenomenon shaped not only by technological performance but also by concerns regarding privacy, ethical considerations, and perceptions of the appropriate role of AI in educational environments.

The findings reveal that students are more likely to distrust AI when they perceive potential risks related to the protection of personal data, question the reliability and accuracy of AI-generated outputs, and believe that AI may replace essential human roles in teaching and learning. These concerns collectively influence students' willingness to engage with AI-assisted learning technologies. Therefore, increasing AI adoption in higher education requires more than technological advancement; it also requires transparent governance, responsible data management, reliable system performance, and the preservation of meaningful human interaction throughout the learning process.

Overall, this study demonstrates that understanding the determinants of AI distrust is essential for promoting responsible and sustainable AI implementation in higher education. Addressing students' concerns regarding privacy, system performance, and the role of human educators will be critical to fostering greater confidence in AI-assisted innovation learning and ensuring that artificial intelligence serves as a valuable complement to, rather than a replacement for, human intelligence and educational expertise.

From a practical perspective, the findings offer several important implications for higher education institutions, educators, and AI developers. Universities should establish clear policies regarding ethical AI use, data privacy, and responsible AI governance to strengthen students' confidence in AI-assisted learning. Educators should encourage students to use AI critically as a complementary learning tool while maintaining independent thinking, creativity, and academic integrity. Furthermore, AI developers should continue improving system accuracy, transparency, explainability, and privacy protection to reduce users' concerns and foster greater trust in AI technologies.

Despite these contributions, this study has several limitations. First, the research focused exclusively on university students, limiting the generalizability of the findings to other populations such as educators, researchers, or professionals. Second, the study examined only three determinants of AI distrust, whereas other factors including AI literacy, algorithm transparency, perceived usefulness, institutional trust, ethical awareness, and previous experience with AI may also influence students' perceptions. Future research is encouraged to incorporate additional psychological, technological, and organizational variables to develop a more comprehensive model of AI distrust.

5. REFERENCES

Chalmers, D., MacKenzie, N. G., & Carter, S. (2021). Artificial intelligence and entrepreneurship: Implications for venture creation in the fourth industrial revolution. Entrepreneurship Theory and Practice, 45(5), 1028–1053. https://doi.org/10.1177/1042258720934581

Chen, L., Ifenthaler, D., Yau, J. Y. K., & Sun, W. (2024). Artificial intelligence in entrepreneurship education: A scoping review. Education + Training, 66(6), 589–608. https://doi.org/10.1108/ET-05-2023-0169

Dwork, C., & Minow, M. (2022). Distrust of artificial intelligence: Sources and responses from computer science and law. Dædalus, 151(2), 309–321. https://doi.org/10.1162/daed_a_01918

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). "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. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Gerlich, M. (2024). Exploring motivators for trust in the dichotomy of human–AI trust dynamics. Social Sciences, 13(5), 251. https://doi.org/10.3390/socsci13050251

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). A primer on partial least squares structural equation modeling (PLS-SEM) (3rd ed.). SAGE Publications.

Kumari, S. (2026). *Disengagement drivers in AI-enabled financial service advisors: The mediating role of distrust.*

Panahi, S., Bazrafshani, A., & Mirzaie, A. (2023). Development and validation of a modified LibQUAL scale in health sciences libraries: Application of structural equation modeling. Journal of the Medical Library Association, 111(4), 792–801. https://doi.org/10.5195/jmla.2023.1348

Pelau, C., Dabija, D. C., & Ene, I. (2021). What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry. Computers in Human Behavior, 122, 106855. https://doi.org/10.1016/j.chb.2021.106855

Sarstedt, M., Hair, J. F., Ringle, C. M., & Richter, N. F. (2024). Partial least squares structural equation modeling (PLS-SEM): A comprehensive guide. Springer.

Smirnova, K., Vecchiarini, M., & colleagues. (2026). *Student distrust of AI in entrepreneurship education contexts: Development and validation of an empirical scale*. *Journal of Small Business and Enterprise Development.*

Tiwari, C. K., Bhat, M. A., Khan, S. T., Subramaniam, R., & Khan, M. A. I. (2024). What drives students toward ChatGPT? An investigation of the factors influencing adoption and usage of ChatGPT. Interactive Technology and Smart Education, 21(3), 333–355. https://doi.org/10.1108/ITSE-04-2023-0061

Vaassen, B. (2022). AI, opacity, and personal autonomy. Philosophy & Technology, 35(4), 88. https://doi.org/10.1007/s13347-022-00577-5