JAAIE : Journal of Applied Artificial Intelligence in Education

Journal of Applied Artificial Intelligence in Education
E-ISSN: 3109-7081
https://journal.lontaradigitech.com/JAAIE

AI Hallucinations in AIED and Their Impact on Students' Intentions to Behave Honestly: A PLS-SEM Analysis of JTIK UNM Students

Desitha Cahya[1*], Putri Ramdani[2], , Annajmi Rauf[3], Andi Baso Kaswar[4], M Miftach Fakhri[5]
[1,2,3] *Universitas Negeri Makassar, Indonesia*
[4] *Okayama University, Japan*
[5] *Universitas Pendidikan Indonesia, Indonesia*

Corresponding e-mail : desithachy@gmail.com

ARTICLE INFO

Keywords:

Algorithmic Bias;
Artificial Intelligence in;
Education;
Digital Literacy;
Honest Behaviour;
Transparency

Article History

Received: Oct 20, 2025
Revised : Nov 30. 2025
Accepted : Jan 5, 2026

ABSTRACT

Artificial Intelligence in Education (AIED) is increasingly used to support learning efficiency, personalization, and academic productivity. However, issues such as AI hallucination, algorithmic bias, limited system Transparency, and variations in students' Digital Literacy present ethical risks that may undermine academic integrity. These challenges indicate a gap between the ideal function of AI as a learning assistant and its practical use, which remains prone to plagiarism and misuse. This study aims to analyze how students' perceptions of algorithmic bias, Transparency in AI systems, and Digital Literacy influence their Honest Behavior when using AI for academic purposes. A quantitative research method was employed using a survey design, and data were analyzed through Partial Least Squares Structural Equation Modeling to empirically examine the relationships among variables. The results show that algorithmic bias, Transparency, and Digital Literacy each have a positive effect on honest behavior, with Digital Literacy emerging as the strongest predictor. These findings suggest that students with better digital skills and awareness of AI mechanisms are more capable of using AI responsibly and ethically. This study concludes that higher education institutions need to strengthen policies related to ethical AI use and enhance students' Digital Literacy to foster an academically honest environment. The study contributes to the development of ethical behavior frameworks in the AIED context and provides considerations for institutions to improve integrity in AI-assisted learning.

To cite this article : Author. (2026). Title. Journal of Applied Artificial Intelligence in Education, 1(2), 49-62.

INTRODUCTION

The development of Artificial Intelligence in Education (AIED) has brought about major changes in the learning process and academic [1], [2]. Although it has the potential to improve learning efficiency and personalization, AIED also presents various challenges. From a technical perspective, issues such as AI hallucinations, algorithmic bias, and data privacy have arisen [3]. Meanwhile, ethical aspects and human values demand fairness, transparency, and accountability in the use of technology [4]. In addition, students' digital access and literacy readiness determine their ability to use AI responsibly [5]. These factors are thought to influence students' intentions to behave honestly when using academic AI.

The use of AI in academic activities is currently increasing, especially among Indonesian students. Most students have utilized ChatGPT and similar platforms to assist with assignment preparation [5]. However, many remain unaware of the risks of errors and potential plagiarism that may result from such use. This situation highlights the gap between the intended use of AI to support academic integrity and practices in the field that remain vulnerable to misuse of technology [1]. In addition, technical challenges such as AI hallucinations that generate false citations can reduce the reliability and academic integrity of students [3], [4].

Technical challenges such as AI hallucination in citation results and algorithmic bias can reduce the reliability of scientific sources generated by AI systems [3]. On the other hand, ethical issues and human values in the use of AI require Transparency, accountability, and moral awareness on the part of users in order to maintain academic integrity [4]. Students' Digital Literacy and readiness also play an important role in determining how AI technology is used responsibly [2], [5].

Although various studies have discussed AIED, most are still conceptual and focus on primary and secondary education; therefore empirical evidence in the context of Indonesian students is still limited [1], [2]. Several studies have highlighted technical challenges such as AI hallucination and algorithmic bias [3], however, it has not yet linked this to ethical aspects such as Transparency and its impact on students' honest behavior. In addition, students' Digital Literacy is rarely measured quantitatively, even though it is an important factor in the responsible use of academic AI [5]. Thus, there is still a research gap in simultaneously examining the influence of Algorithm Bias, Transparency, and Digital Literacy on students' Honest Behavior in the context of academic AI use.

Theoretically, this study is based on the Theory of Planned Behavior (TPB) developed by Icek Ajzen to explain and predict behavior through attitude, subjective norms, and perceived behavioral control [6]. In the context of AIED, algorithmic bias and transparency can influence students' attitudes and trust in the reliability of AI systems, while Digital Literacy plays a role in strengthening perceptions of behavioral control in maintaining academic integrity. TPB explains that a person's intention to act honestly arises when they have positive attitudes, social support, and perceptions of their ability to control their actions [6]. This framework is consistent with the need to understand the Honest Behavior of JTIK UNM students in facing the technical and ethical challenges of academic AI use.

This research is important because it integrates three main aspects of AIED, namely algorithmic bias, Transparency, and Digital Literacy, which are interrelated in shaping students' honest behavior. These findings are expected to provide practical contributions to universities in formulating ethical policies on the use of academic AI that promote integrity and digital responsibility. In the national context, the increased use of AI by students without adequate understanding raises the need to strengthen technological literacy and internal campus regulations. This research is also relevant to current issues regarding AI hallucination and data reliability in academic processes, which are of global concern in higher education.

This study aims to analyze the influence of Algorithm Bias, Transparency, and *Dig*ital Litera*cy* on students' Honest Behavior in the use of academic AI. The first specific objective is to determine the extent to which Algorithm Bias affects the Honest Behavior of JTIK UNM students. The second objective is to analyze the influence of AI system Transparency on students' tendency to maintain academic honesty. The third objective was to measure students' Digital Literacy levels and their relationship with Honest Behavior in the use of AI technology. The results of this study are expected to provide empirical contributions to the development of digital ethical behavior theory as well as form the basis for the formulation of campus policies oriented towards academic integrity and responsible use of AI.

RQ 1: Does Algorithm Bias have a positive and significant effect on Honest Behavior?

RQ 2: Does Transparency have a positive and significant effect on Honest Behavior?

RQ 3: Does Digital Literacy have a positive and significant effect on Honest Behavior?

METHOD

Research Design

This study used a quantitative approach with a cross-sectional design to analyze the influence of algorithm bias, transparency, and digital literacy on students' honest behavior in using Artificial Intelligence in Education (AIED) systems in an academic environment. A quantitative approach was chosen because it allows for empirical and objective measurement of the relationship between variables based on numerical data [7]. The cross-sectional design was chosen because data was collected in a single period of time to describe the actual conditions of students without direct intervention from researchers [8]. This research design is in line with the objective of understanding the phenomenon of students' ethical behavior towards AI technology, particularly in the context of technical and ethical challenges such as AI hallucination and algorithmic bias that can affect academic integrity. This model is also relevant because it is able to represent students' perceptions of academic AI systems in a real context in higher education [4], [9].

Participant

The research participants consisted of 92 active students from the Department of Informatics and Computer Engineering (JTIK) at Universitas Negeri Makassar who were directly involved in artificial intelligence-based learning activities (Artificial Intelligence in Education/AIED). The criteria for selecting participants included: (1) Students who were active in the 2022–2025 academic year, (2) Had experience using generative AI platforms for academic purposes, (3) Owned personal digital devices, and (4) Understood the basics of digital ethics. This selection ensured that the sample was relevant to the research context, namely students' ethical behavior in the use of academic AI.

To ensure that the number of participants was adequate and in line with the needs of the research analysis, the sample size was determined through rigorous methodological considerations. The sample size was determined by considering the complexity of the model and the number of indicators in the study, so that the number was deemed adequate for the analysis. This approach ensures that the sample size is sufficient to produce valid research results that can be generalized to the broader population of UNM students. In addition, this number also meets the minimum limit for analysis using Partial Least Squares–Structural Equation Modeling (PLS-SEM), where a sample size of around 90 respondents is considered sufficient for a model with three independent variables and moderate effect strength [10].

Population and the methods of sampling Instrumentation

The population of this study includes all active UNM students involved in AI-based learning. The selection of this population is based on the increasing use of platforms such as ChatGPT and Gemini in the learning process of students. The use of this technology has proven to be part of students' academic activities, particularly in digital literacy and academic task completion [10]. This condition shows that UNM students already have direct experience with the use of AI in an educational context. Therefore, this population is considered relevant to support the research objectives.

In addition to the relevance of AI use, UNM JTIK students also showed a high level of technology adoption in supporting their learning process. Activities such as completing assignments, searching for references, and academic analysis were often carried out with the help of generative AI platforms. This is in line with previous research findings which show that students in the field of technology tend to be more active in utilizing ChatGPT and similar applications for academic needs [11]. The intensity of this utilization reinforces the rationale for selecting JTIK students as part of the research population. Thus, this group serves as an appropriate representation for examining how AI is used in an academic context.

The sampling technique used was purposive sampling, a non-probabilistic technique that sets specific criteria for selecting respondents [12]. This technique was chosen because it was able to identify students who truly had relevant experience in using AI technology. Previous studies have shown that academic studies on AI require respondents who understand digital ethics and have had direct interaction with AI applications [13]. Therefore, purposive sampling is considered the most appropriate method to describe the actual behavior of students who use AI. This technique ensures that only individuals who meet certain characteristics are recruited as research samples.

Instrument

Data collection in this study was conducted using a questionnaire in the form of a Google form that was distributed online via WhatsApp to active students at Universitas Negeri Makassar. The instrument was used to facilitate respondent access, expand the range of participants, and adapt to the characteristics of students who are active in digital learning [5]. The research instrument was a structured questionnaire based on four research variables, namely Algorithmic Bias (AB), Transparency (T), Digital Literacy (DL), and Honest Behavior (HB). Each variable had five statement items designed to measure students' perceptions of technical, ethical, and honest behavior challenges in the use of academic AI.

The content and formulation of statements in the questionnaire have been verified through content validity by expert judgment to ensure clarity of language, relevance of indicators, and suitability for the research objectives. Content validity was conducted to assess the extent to which each item truly represents the construct being measured, so that the instrument has conceptual suitability with the research variables [7]. This survey uses a five-point Likert scale for all statement items, with a range of answers from (1 = Strongly Disagree) to (5 = Strongly Agree). This scale was chosen so that respondents could express their level of agreement clearly, measurably, and consistently.

Table 1. Research Instruments

No.	Variable	Statement	Reference
1	Algorithmic Bias (AB)	1, 2, 5	[4]
2	Transparency (TP)	1, 3, 5	[14], [15]
3	Digital Literacy (DL)	2, 3, 5	[16], [17]
4	Honest Behavior (HB)	2, 4, 5	[18]

Procedures

This research procedure began with the identification of issues related to the challenges of Artificial Intelligence in Education (AIED), followed by the formulation of problems based on the gaps found in the context of AI use by students. Next, a literature review was conducted covering seven scientific articles to strengthen the theoretical basis and direct the focus of the research. Based on the results of this review, the researchers developed hypotheses, determined the population and sample, and selected purposive sampling techniques in accordance with the research objectives. After that, a research instrument in the form of a questionnaire was developed and content validation was carried out by expert judgment before the instrument was converted into a Google Form. The next stage was data collection and analysis using PLS-SEM to evaluate the measurement model and structural model. The results of the analysis were then used to interpret the findings, which were ultimately summarized in the conclusion and recommendations as the final stage of the research.

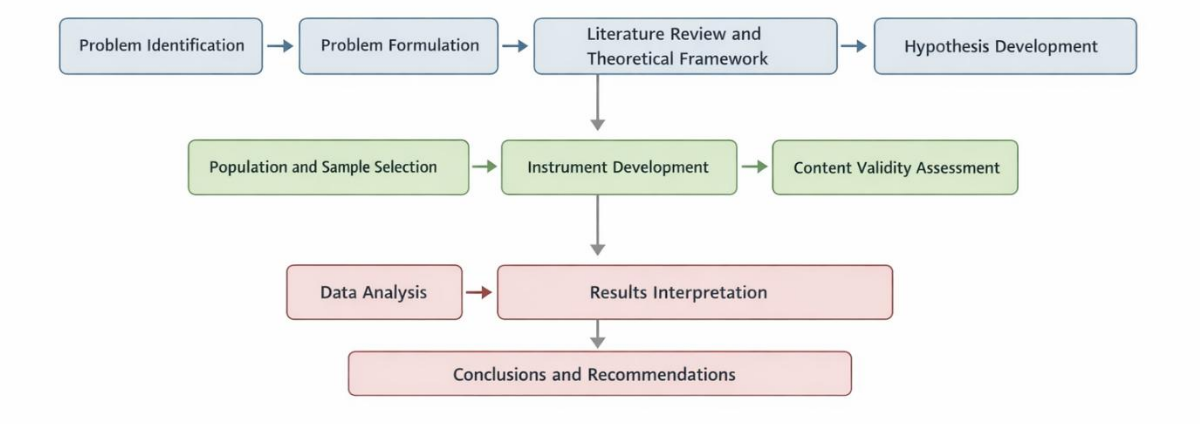

Figure 1 Research Procedure Flowchart

Analysis Plan

This study used the Partial Least Squares Structural Equation Modeling (PLS-SEM) method, which is a multivariate statistical technique for analyzing complex relationships between latent variables and their indicators [8]. Unlike covariance-based SEM, PLS-SEM focuses on maximizing the variance explained by the model, making it suitable for exploratory research with small samples and non-normally distributed data This method is relevant for examining the interrelationships between variables in the use of AIED technology

The analysis was conducted using SmartPLS 4 to test the measurement model (Outer Model) including convergent validity, discriminant validity, and construct reliability as well as the structural model to assess the relationship between latent variables through path coefficients [19]. This stage ensures that the model is able to accurately describe the contribution of each variable.

In the context of this study, PLS-SEM was used to examine the influence of Algorithmic Bias (X1), Transparency (X2), and Digital Literacy (X3) on Honest Behavior (Y). This approach was chosen because it is capable of explaining complex causal relationships that are relevant to technical and ethical issues in AIED. In addition, descriptive statistical analysis was used to describe the demographic characteristics of respondents, including age, gender, education level, and experience using AI-based technology.

Outer Model

Outer model evaluation is conducted to ensure that the indicators in the Algorithm Bias, Transparency, Digital Literacy, and Honest Behavior constructs accurately represent the latent variables before being analyzed in the structural model. This stage is very important in PLS-SEM because latent variables cannot be measured directly, so the quality of the indicators determines the accuracy of the constructs. This test also ensures the consistency of the indicators and guarantees that the measurement model meets the instrument quality standards before structural analysis is performed. Thus, validity and reliability tests are necessary to ensure that the indicators accurately describe the constructs [20].

In accordance with PLS-SEM guidelines, the evaluation of the outer model includes convergent validity, discriminant validity, and construct reliability [21]. Convergent validity has been fulfilled, as indicated by outer loading values ≥ 0.708 and AVE ≥ 0.50 for all indicators, which signifies the ability of the indicators to adequately explain the latent variables. These results form the basis for continuing the evaluation of other validities and strengthening the reliability of the research instrument [21].

Discriminant validity was also verified using the Fornell–Larcker criteria, whereby the square root of the AVE for each construct was higher than the inter-construct correlation [22]. Furthermore, the reliability of the constructs tested using Composite Reliability and Cronbach's Alpha showed values > 0.70, indicating good internal consistency across all constructs [21]. These results ensure that the indicators work stably and reliably in describing latent variables.

Overall, the fulfillment of convergent validity, discriminant validity, and construct reliability indicates that the research measurement tools have met quality standards. In the initial stage, indicators that did not meet the criteria were eliminated so that the measurement model was more accurate. The remaining indicators were assessed as representative and stable in various tests, making them suitable for use in structural analysis and capable of producing scientifically accountable findings.

Inner Model

Next, the second stage is the evaluation of the inner model to assess the causal relationships between latent variables in the structural model. This evaluation includes path coefficient analysis to see the direction and strength of the relationships, as well as significance testing through t-statistics and p-values obtained through the bootstrapping procedure. The inner model in PLS-SEM allows researchers to understand the simultaneous contributions of Algorithmic Bias, Transparency, and Digital Literacy in predicting students' Honest Behavior.

Overall, the use of PLS-SEM provides methodological advantages in comprehensively describing the relationships between variables, so that the results of this study can provide a deeper understanding of the factors that influence students' honest behavior in the use of AI technology in an academic context.

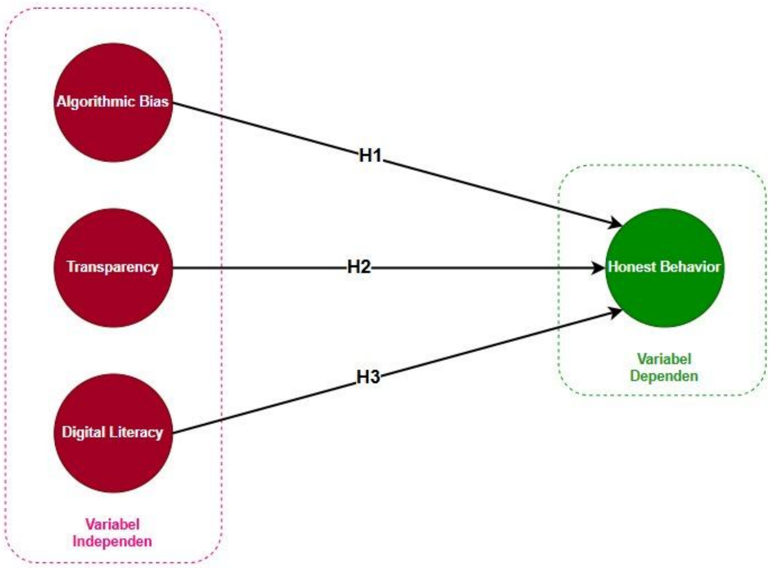

Figure 2. The model proposed in this study

Hipotesis:

H1: Algorithm Bias *berpengaruh positif dan signifikan terhadap* Honest Behavior.

H2: Transparency *berpengaruh positif dan signifikan terhadap* Honest Behavior.

H3: Digital Literacy *berpengaruh positif dan signifikan terhadap* Honest Behavior.

RESULTS AND DISCUSSION

Results

Demographic Respondents

A total of 97 respondents participated in this study. To provide a clearer understanding of the respondent profile, demographic data is presented in Table 2 below. This information covers several important aspects, namely gender, age range of respondents, study program, current semester, class year, ownership of digital devices, and frequency of technology use for learning activities.

Table 2. Respondent Demographic Data

No.	Category	Description	Percentage
1.	Gender	Male	42.3%
		Female	57.7%
2.	Age	17	3.1%
		18	23.7%
		19	47.4%
		20	19.6%
		21	6.2%
3.	Study Program	PTIK	80.4%
		TEKOM	19.6%
4.	Vacation	I	23.7%
		III	64.9%
		V	8.2%
		VII	3.1%
4.	Generation	2022	3.1%
		2023	9.3%
		2024	63.9%
		2025	23.7%
5.	Ownership of digital devices (laptops, tablets, or smartphones)	Yes	100.0%
		No	-

6.	Frequency of Technology Use for Learning	1–2 times a week	3.1%
		3–5 times a week	15.5%
		Every day	81.4%

Based on Table 2 above, the gender distribution shows that female respondents are slightly more dominant (57.7%) than male respondents (42.3%). In terms of age, most respondents were in the 19-year-old age group (47.4%), followed by 18-year-olds (23.7%) and 20-year-olds (19.6%). In terms of academic background, the majority of respondents were third-semester students (64.9%), while first-semester (23.7%) and fifth-semester (8.2%) students were fewer in number. Based on their class year, the largest group came from the class of 2024 (63.9%), followed by the class of 2025 (23.7%). In terms of study programs, respondents were predominantly PTIK students (80.4%), while TEKOM contributed (19.6%). All respondents (100%) had digital devices. The majority used technology 3–5 times per week (81.4%). Overall, respondents were young students familiar with digital technology.

Outer Model

Convergent Validity and Construct Reliability

Table 3 presents the results of convergent validity and construct reliability testing for the four constructs: Algorithm Bias (AB), Transparency (TP), Digital Literacy (DL), and Honest Behavio*r* (HB). All constructs meet the recommended thresholds for PLS-SEM analysis, indicating good measurement quality.

Table 3. Results of Convergent Validity and Construct Reliability Evaluation

Construct and Items	Outer Loading	Rho_A	Composite Reliability (CR)	Average Variance Extracted (AVE)
Algorithm Bias (AB)				
AB1	0.791	0.815	0.886	0.723
AB2	0.890			
AB5	0.867			
Digital Literacy (DL)				
DL2	0.909	0.911	0.943	0.846
DL3	0.930			
DL5	0.921			
Honest Behavior (HB)				
HB2	0.922	0.891	0.932	0.820
HB4	0.883			
HB5	0.911			
Transparency (TP)				
TP1	0.774	0.815	0.884	0.718
TP3	0.868			
TP5	0.896			

In the Algorithm Bias construct, the outer loading values range from 0.791 to 0.890, with Rho_A of 0.815, Composite Reliability (CR) of 0.886, and Average Variance Extracted (AVE) of 0.723. These values confirm that the indicators consistently represent the AB construct. The Transparency construct also demonstrates strong measurement performance, with loadings between 0.774 and 0.896, Rho_A of 0.815, CR of 0.884, and AVE of 0.718.

The Digital Literacy construct shows excellent convergent validity, with outer loading values ranging from 0.909 to 0.930, Rho_A of 0.911, CR of 0.943, and AVE of 0.846. Meanwhile, Honest Behavior has loading values between 0.883 and 0.922, Rho_A of 0.891, CR of 0.932, and AVE of 0.820, indicating strong internal consistency

Overall, all constructs exceed the minimum criteria recommended by [23], namely outer loading > 0.70, CR and Rho_A > 0.70, and AVE > 0.50. Thus, the measurement model fulfills the convergent validity and reliability requirements and is suitable for further analysis.

Discriminant Validity

Table 4 below presents the results of discriminant validity testing using the Fornell-Larcker criteria for four research constructs, namely Algorithm Bias, Transparency, Digital Literacy, and Honest Behavior. This test was conducted by comparing the AVE square root value on the diagonal with the correlation between constructs in rows and columns. A construct is declared to meet discriminant validity if the AVE square root value is higher than the correlation of other constructs, thus indicating that each construct is able to distinguish itself from other constructs in the model.

Table 4. Results of the Fornell-Lacker Validity Test

	Algorithm Bias	Digital Literacy	Honest Behavior	Transparency
Algorithm Bias	0.850			
Digital Literacy	0.703	0.920		
Honest Behavior	0.741	0.796	0.906	
Transparency	0.750	0.649	0.699	0.848

Based on the test results in Table 4 above, all constructs meet the Fornell–Larcker criterion. The Algorithm Bias construct has an AVE square root value of 0.850, which is higher than its correlations with Digital Literacy (0.703), *Honest Behavior* (0.741), and Transparency (0.750). This indicates that the *Algorithm Bias* construct is clearly distinguishable from other constructs.

The Digital Literacy construct also meets the discriminant validity requirement, as its AVE square root (0.920) exceeds its correlations with Algorithm Bia*s* (0.703), Honest Behavior (0.796), and Transparency (0.649). Similarly, the Honest Behavior construct has an AVE square root value of 0.906, higher than its correlations with Algorithm Bias (0.741), Digital Literacy (0.796), and Transparency (0.699).

The *Transparency* construct shows an AVE square root of 0.848, which is greater than its correlations with Algorithm Bias (0.750), Digital Literacy (0.649), and Honest Behavior (0.699). These findings confirm that each construct in the model is conceptually distinct and does not overlap excessively with others. Therefore, the research model meets the discriminant validity requirements and is appropriate for continued structural analysis.

Inner Model

To test the relationship between variables in the research model, an analysis was conducted using the PLS-SEM approach. This test aimed to examine the influence of Algorithm Bias, Transparency, and Digital Literacy on Honest Behavior. The following table presents the complete results of the hypothesis test, including path coefficient, t-statistic, p-value, and significance decision.

Table 5. Test Results of the Relationship between Latent Constructs

	Hypothesis	Path Coefficient	T-statistics	P-values	Decision
H1	Algorithm Bias -> Honest Behavior	0.248	2.420	0.008	Positive and significant
H2	Transparency -> Honest Behavior	0.188	1.920	0.000	Positive and significant
H3	Digital Literacy -> Honest Behavior	0.499	5.457	0.027	Positive and significant

Based on the results presented in Table 5, all hypotheses in the structural model show significant effects, demonstrating that Algorithm Bias, Transparency, and Digital Literacy each contribute to explaining Honest Behavior. *Algorithm Bias* has a positive and significant effect on Honest Behavior (β = 0.248; t = 2.420; p = 0.008), indicating that students who recognize potential bias in AI systems tend to be more cautious and behave honestly when utilizing academic AI tools.

Transparency also exhibits a positive and significant influence on Honest Behavior (β = 0.188; t = 1.920; p = 0.000). This finding suggests that when AI systems provide clearer information, more explainable outputs, and transparent mechanisms, students are more likely to use them responsibly and maintain academic integrity.

Digital Literacy shows the strongest effect on Honest Behavior (β = 0.499; t = 5.457; p = 0.027). This result highlights that students with higher levels of digital skills have a better understanding of how AI systems operate, are more capable of evaluating the accuracy of AI-generated information, and demonstrate greater ethical awareness in academic activities.

Discussion

The first hypothesis, Algorithm Bias Behavior, shows that Algorithm Bias has a positive and significant effect on Honest Behavior, with a path coefficient of 0.248, a t-value of 2.420, and a p-value of 0.008, thus H1 is accepted. This finding is in line with research showing that bias in algorithms or AI-based recommendations can influence a person's tendency to be honest or otherwise, where more neutral and unbiased systems tend to encourage more ethical behavior [24]. These results reinforce the idea that students who recognize potential bias in AI outputs tend to be more careful, reflective, and ethically responsible in evaluating AI-generated information.

The second hypothesis, Transparency Behavior, shows that this hypothesis has a positive and significant effect on Honest Behavior, with a path coefficient of 0.188, a t-value of 1.920, and a p-value of 0.000. This result aligns with studies explaining that Transparency in algorithmic decision-making mechanisms, such as revealing reasons, features, or weighted factors behind decisions, can reduce suspicion, build perceptions of fairness, and increase users' tendency to be honest [25]. Greater Transparency creates a more ethical and accountable environment [25], suggesting that clear AI system explanations foster responsible academic behavior.

Furthermore, the third hypothesis, Digital Literacy Behavior, shows that Digital Literacy has a positive and significant effect on Honest Behavior, with a path coefficient of 0.499, a t-value of 5.457, and a p-value of 0.027, thus accepting H3. These results confirm that Digital Literacy is the strongest factor in encouraging honest behavior. Emphasizes that individuals with high Digital Literacy are better able to understand how algorithms work, assess the reliability of information, and make ethical decisions based on that understanding [26]. This implies an increase in Honest Behavior because users can distinguish between valid and invalid information [26]. These findings also correspond with the idea that understanding bias, fairness, and algorithmic processes in digital systems encourages more responsible and ethical conduct in digital environments [27].

Overall, the results of this study indicate that the three independent variables Algorithm Bias, Transparency, and Digital Literacy, play an important role in shaping Honest Behavior. An unbiased system, transparent processes, and strong digital skills reinforce individuals' tendency to act honestly in a digital context. These findings support previous research showing that technological factors and user literacy work simultaneously in shaping ethical behavior in the digital age.

This study makes an important contribution to understanding the factors that influence students' Honest Behavior in the context of AI use in academic settings. Findings regarding the role of Algorithm Bias, Transparency, and Digital Literacy enrich the AIED literature by confirming that technical aspects and digital competencies directly contribute to users' ethical behavior. Furthermore, this study offers an empirical model based on PLS-SEM that can be used as a foundation for future research exploring ethical aspects of AI in education. This model can also guide educational institutions in designing AI usage policies that are fairer, more transparent, and oriented toward academic integrity. Thus, this study not only offers empirical results but also provides practical contributions to higher education.

Although this study provides significant findings, there are several limitations to consider. First, the study only involved students from one faculty at one institution, limiting the generalizability of the results. Second, the data were collected through perception-based questionnaires, making the findings dependent on respondent honesty and subjectivity. Third, this study only tested three independent variables, while other important factors such as trust in AI, perceived fairness, or ethical awareness were not included. In addition, the cross-sectional research design did not allow the observation of behavioral changes over time.

Further research should expand the respondent population to include multiple faculties or universities to increase generalizability. Additional variables, such as trust in AI, AI self-efficacy, or ethical awareness, may help enrich the research model. Longitudinal approaches are recommended to observe how ethical behavior evolves with changes in AI usage patterns over time. A mixed methods approach can also be employed to explore students' experiences more deeply so that quantitative findings can be strengthened with qualitative insights. With such developments, research in the field of AIED is expected to become more comprehensive and applicable.

CONCLUSIONS

The findings of this study confirm that the expectations formulated in the Introduction, namely the influence of Algorithmic Bias, Transparency, and Digital Literacy on students' Honest Behavior in the context of AIED, are aligned with the empirical evidence presented in the Results and Discussion. All three variables show positive and significant effects, with Digital Literacy emerging as the strongest predictor, demonstrating that ethical behavior in AI-assisted academic activities is shaped by both technological factors and students' digital competencies.

This research contributes theoretically by strengthening the application of the Theory of Planned Behavior (TPB) in the AIED context, showing that attitudes shaped by algorithmic bias, perceived transparency, and perceived behavioral control through digital literacy significantly influence honest behavior. Methodologically, the study provides an empirical model using PLS-SEM that can be replicated or expanded by future researchers to explore ethical decision-making in AI-supported learning. Practically, the results offer institutions a data-driven foundation for designing policies on ethical AI use, emphasizing the need for transparent AI systems and strengthened digital literacy training to promote academic integrity.

The limitations of this study, such as the focus on a single faculty, reliance on self-reported perceptions, and the use of a cross-sectional design, restrict the generalizability of the findings and may influence the stability of behavioral estimates over time. These limitations imply that the results should be interpreted with caution when applied to broader populations or different academic settings.

Future research should involve larger and more diverse samples across multiple faculties or universities to improve generalizability. Additional variables such as trust in AI, perceived fairness, AI self-efficacy, or ethical awareness could enhance the explanatory power of future models. Longitudinal designs are recommended to observe how honest behavior evolves as students' exposure to AI increases. Further studies may also explore practical implementations, such as integrating AI literacy modules or evaluating AI transparency interventions to strengthen ethical behavior in academic environments. Through such developments, the prospects for research and application in AIED can continue to progress and support responsible and integrity-based AI usage in higher education.

REFERENCES

[1] S. Budiyono, P. Azhari, and M. A. B. Pamungkas, "Problem Penggunaan AI (Artificial Intelligence) dalam Bidang Pendidikan," *Al-DYAS*, vol. 3, no. 2, pp. 660–669, May 2024, doi: 10.58578/aldyas.v3i2.2935.

[2] R. Setiawan and N. Sukmana, "Kalam Cendekia: Jurnal Ilmiah Kependidikan Kebijakan Penggunaan Teknologi Artificial Intelligence (AI) dalam Pembelajaran di Sekolah Dasar," *Kalam Cendekia: Jurnal Ilmiah Kependidikan*, vol. 13, pp. 1137–1143, 2025, doi: 10.20961/jkc.v13i2.102104.

[3] F. Aljamaan *et al.*, "Reference Hallucination Score for Medical Artificial Intelligence Chatbots: Development and Usability Study," *JMIR Med Inform*, vol. 12, 2024, doi: 10.2196/54345.

[4] A. M. Al-Zahrani, "Unveiling the shadows: Beyond the hype of AI in education," *Heliyon*, vol. 10, no. 9, May 2024, doi: 10.1016/j.heliyon.2024.e30696.

[5] A. Supriyono, A. Djoko Lesmono, T. Prihandono, D. dan Tantangan Pemanfaatan ChatGPT dalam Pembelajaran pada Kurikulum Merdeka, T. Literatur Sistematis Jurnal Pendidikan dan Kebudayaan Vol, and A. Supriyono Albertus Djoko Lesmono, "Dampak dan Tantangan Pemanfaatan ChatGPT dalam Pembelajaran pada Kurikulum Merdeka: Tinjauan Literatur Sistematis The Impact and Challenges of Utilizing ChatGPT in Learning within the Kurikulum: A Systematic Literature Review," *Jurnal Pendidikan dan Kebudayaan*, vol. 9, no. 2, 2024, doi: 10.24832/jpnk.v9i2.5214.

[6] I. Ajzen, "The theory of planned behavior: Frequently asked questions," *Hum Behav Emerg Technol*, vol. 2, no. 4, pp. 314–324, Oct. 2020, doi: 10.1002/hbe2.195.

[7] J. W. Creswell and J. D. Creswell, "Research Design Qualitative, Quantitative, and Mixed Methods Approaches Fifth Edition," *Sage Open*, 2018.

[8] J. F. Hair, G. Tomas, M. Hult, C. M. Ringle, and M. Sarstedt, "A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)," 2022. [Online]. Available: https://www.researchgate.net/publication/354331182

[9] M. S. Rahman, M. M. Sabbir, J. Zhang, I. H. Moral, and G. M. S. Hossain, "Examining students' intention to use ChatGPT: Does trust matter?," *Australasian Journal of Educational Technology*, vol. 39, no. 6, pp. 51–71, Dec. 2023, doi: 10.14742/ajet.8956.

[10] D. Aswan, "Hubungan antara literasi digital dan persepsi mahasiswa tentang etika penggunaan AI di kalangan akademik," *Jurnal Ilmiah Wahana Pendidikan*, vol. 11, no. 6.D, pp. 283–294, 2025. [Online]. Available: https://jurnal.peneliti.net/index.php/JIWP/article/view/12159

[11] E. Anisa, M. Ulfa, and M. D. Anggriani, "Dampak penggunaan AI ChatGPT dalam pembelajaran di kalangan mahasiswa PGSD Universitas Riau," *ENTINAS: Jurnal Pendidikan dan Teknologi Pembelajaran*, vol. 3, no. 2, pp. 46–54, 2025.

[12] I. Etikan and K. Bala, "Sampling and Sampling Methods," *Biom Biostat Int J*, vol. 5, no. 6, May 2017, doi: 10.15406/bbij.2017.05.00149.

[13] N. Saputri and S. Surawan, "Interaksi Mahasiswa Dengan Artificial Intelligence dan Implikasinya Terhadap Akhlak Digital: Tinjauan Psikologi Pendidikan Islam," *Al-Zayn : Jurnal Ilmu Sosial & Hukum*, vol. 3, no. 3, pp. 1606–1616, Jun. 2025, doi: 10.61104/alz.v3i3.1467.

[14] L. Yu and Y. Li, "Artificial Intelligence Decision-Making Transparency and Employees' Trust: The Parallel Multiple Mediating Effect of Effectiveness and Discomfort," *Behavioral Sciences*, vol. 12, no. 5, May 2022, doi: 10.3390/bs12050127.

[15] J. Hansson and E. Hubendick, "Public perceptions of transparency in AI-driven decision-making in Sweden-A quantitative analysis of Swedish citizens' views on ethical concerns in public sector AI," 2025.

[16] S. Muammar, P. Maheshwari, and S. Atalla, "An Integrated Theoretical Model for Assessing Digital Literacy's Impact on Academic Performance: A Case Study Using PLS-SEM," *IEEE Access*, vol. 13, pp. 101624–101638, 2025, doi: 10.1109/ACCESS.2025.3578107.

[17] E. Avinç and F. Doğan, "Digital literacy scale: Validity and reliability study with the rasch model," *Educ Inf Technol (Dordr)*, vol. 29, no. 17, pp. 22895–22941, Dec. 2024, doi: 10.1007/s10639-024-12662-7.

[18] M. M. Mohammadi, A. Naghibzadeh, H. Mosafer, and M. Mohammadi, "Measuring honesty in nursing: scale development and validation," *BMC Nurs*, vol. 24, no. 1, Dec. 2025, doi: 10.1186/s12912-025-03163-0.

[19] J. Benitez, J. Henseler, A. Castillo, and F. Schuberth, "How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory IS research," *Information and Management*, vol. 57, no. 2, Mar. 2020, doi: 10.1016/j.im.2019.05.003.

[20] J. F. Hair and M. A. Sabol, "Partial Least Squares Structural Equation Modeling (PLS-SEM): A Rapidly Emerging SEM Alternative," in *International Encyclopedia of Statistical Science*, Springer Berlin Heidelberg, 2025, pp. 1880–1882. doi: 10.1007/978-3-662-69359-9_466.

[21] Y. Haji-Othman and M. S. S. Yusuff, "Assessing Reliability and Validity of Attitude Construct Using Partial Least Squares Structural Equation Modeling (PLS-SEM)," *International Journal of Academic Research in Business and Social Sciences*, vol. 12, no. 5, May 2022, doi: 10.6007/ijarbss/v12-i5/13289.

[22] S. M. Rasoolimanesh, "Discriminant validity assessment in PLS-SEM: A comprehensive composite-based approach," 2022. [Online]. Available: https://www.scriptwarp.com,

[23] J. F. . Hair, G. T. M. . Hult, C. M. . Ringle, and Marko. Sarstedt, *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage, 2017.

[24] M. Leib, N. Köbis, R. M. Rilke, M. Hagens, and B. Irlenbusch, "CORRUPTED BY ALGORITHMS? HOW AI-GENERATED AND HUMAN-WRITTEN ADVICE SHAPE (DIS)HONESTY," *Economic Journal*, vol. 134, no. 658, pp. 766–784, Feb. 2024, doi: 10.1093/ej/uead056.

[25] S. Ebrahimi, E. Abdelhalim, K. Hassanein, and M. Head, "Reducing the incidence of biased algorithmic decisions through feature importance transparency: an empirical study," *European Journal of Information Systems*, vol. 34, no. 4, pp. 636–664, 2025, doi: 10.1080/0960085X.2024.2395531.

[26] E. Gagrčin, T. K. Naab, and M. F. Grub, "Algorithmic media use and algorithm literacy: An integrative literature review," 2024, *SAGE Publications Ltd*. doi: 10.1177/14614448241291137.

[27] P. Panarese, M. M. Grasso, and C. Solinas, "Algorithmic bias, fairness, and inclusivity: a multilevel framework for justice-oriented AI," *AI Soc*, 2025, doi: 10.1007/s00146-025-02451-2.