

AI literacy mediates AI assisted diagnosis participation and critical thinking among medical students under supervision

Received: 27 July 2025

Accepted: 26 February 2026

Cite this article as: Xin, Y., Yan, D., Shuren, L. *et al.* AI literacy mediates AI assisted diagnosis participation and critical thinking among medical students under supervision. *npj Digit. Med.* (2026). <https://doi.org/10.1038/s41746-026-02521-9>

Yang Xin, Deng Yan, Luo Shuren, Luo Minyang & Lu Liuheng

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

AI literacy mediates AI assisted diagnosis participation and critical thinking among medical students under supervision

Yang Xin ^{1,2}, Deng Yan ³, Luo Shuren ⁴, Luo Minyang ⁵, and Lu Liheng ^{1,*}

¹ Guangxi Orthopedic Hospital, Nanning, 530012, China

² Guangxi University, Nanning, 530004, China, 1246946636@qq.com

³ Guangxi Medical University, Nanning, 530021, China, 641578915@qq.com

⁴ Health Commission of Guangxi Zhuang Autonomous Region, Nanning, 530012, China, luoshuren1023@163.com.

⁵ The Second Affiliated Hospital of Guangxi Medical University, Nanning, 530007, Guangxi, China, lmy362331@163.com

* Corresponding Author, Email: luliheng809@163.com.

Abstract

Concerns that AI tools may erode diagnostic reasoning contrast with claims that AI can foster higher order thinking. This longitudinal study followed 372 medical students across 12 months of supervised rotations using an AI-assisted diagnosis system. AI-assisted diagnosis participation, AI literacy and medical critical thinking were assessed at baseline, 6 months and 12 months. Cross lagged panel models examined prospective associations, statistical mediation by AI literacy and moderation by prior technological experience and learning goal orientation. Higher participation was associated with increases in AI literacy and critical thinking, and AI literacy statistically mediated the participation to critical thinking association. Indirect effects were stronger among students with greater technological experience and mastery oriented goals and weaker among performance oriented peers. Findings indicate that, within supervised clinical training, engagement with AI systems is associated with critical thinking development partly through enhanced AI literacy, supporting AI tools as educational resources under faculty guidance.

Keywords: medical education, artificial intelligence, critical thinking, AI literacy, AI-assisted diagnosis

Introduction

The integration of artificial intelligence (AI) into clinical workflows has transitioned from theoretical potential to daily practice, fundamentally altering the educational landscape for medical trainees [1-3]. Medical students in clinical rotations increasingly engage with AI-driven diagnostic support systems, ranging from radiology interpretation algorithms to large language models (LLMs) assisting in differential diagnosis [4]. Unlike traditional passive learning tools, these systems function as advanced probabilistic support tools, offering diagnostic probabilities and reasoning pathways that students must actively evaluate [5-6]. As AI becomes ubiquitous in medical training, a critical pedagogical question emerges: *does frequent participation in AI-assisted diagnosis enhance medical students' critical thinking capabilities, or does it lead to cognitive offloading and dependency?*

Recent research presents a divided perspective on this cognitive impact. One strand of literature warns of "cognitive atrophy," suggesting that over-reliance on automated diagnostic outputs may erode foundational reasoning skills and foster passivity [7-8]. Conversely, proponents argue that AI tools can serve as "cognitive scaffolds," handling routine information processing and freeing cognitive resources for higher-order analysis and synthesis [9]. However, much of the existing evidence relies on cross-sectional surveys or experimental snapshots, failing to capture the developmental trajectory of cognitive skills over time [10]. Consequently, a substantive gap remains in understanding the longitudinal mechanisms through which sustained interaction with AI systems influences the maturation of critical thinking during clinical training.

To address this gap, the present study frames the relationship between AI usage and critical thinking through the lens of Situated Learning Theory. We posit that "AI-Assisted Diagnosis Participation" within a supervised hierarchical medical team represents not merely a background variable but a form of legitimate peripheral participation—an active learning behaviour embedded in authentic clinical practice. Within this supervisory structure, guidance from senior physicians provides the necessary pedagogical scaffolding for students to interrogate AI outputs without compromising patient safety. Through repeated, situated engagement with AI tools—a process of "learning by doing"—students progressively acquire specific technical and cognitive competencies, herein conceptualised as "AI Literacy" [3,11]. Crucially, we distinguish this construct from mere usage frequency; AI literacy represents the capacity to comprehend algorithmic logic, evaluate probabilistic outputs, and effectively integrate AI recommendations into clinical reasoning [12]. Accordingly, we hypothesise that participation serves as the behavioural antecedent that fosters the developmental acquisition of this sophisticated literacy over time.

Furthermore, this study proposes that AI literacy acts as the essential cognitive bridge to critical thinking development. Medical critical thinking involves the disciplined evaluation of evidence, identification of biases, and synthesis of conflicting data [13-14]. We argue that students with high AI literacy possess the requisite "algorithmic awareness" to interrogate AI-generated suggestions rather than accepting them blindly. This evaluative process—identifying algorithmic hallucinations, recognizing data biases, and reconciling AI outputs with clinical context—

constitutes a rigorous exercise in critical reasoning [15-16]. Thus, AI literacy theoretically mediates the relationship between participation and critical thinking: active usage builds the competence (literacy) necessary to engage in the high-level evaluation (critical thinking) that prevents cognitive dependence.

Recognizing that this developmental process is not uniform, the analyses also incorporate individual difference factors based on Achievement Goal Theory [17]. We recognize that the developmental pathways linking AI-assisted diagnosis participation, AI literacy, and medical critical thinking are likely to vary across learners [5]. Prior technological experience may shape how comfortable students feel experimenting with AI tools, while learning goal orientations may influence whether they approach AI-assisted tasks as opportunities to deepen understanding or primarily as means to secure correct answers [18,19,20]. Mastery-oriented students, who prioritize learning and competence development, may be more inclined to use AI tools for exploration and reflection, whereas performance-oriented students may focus on efficiency and error avoidance, potentially reinforcing shortcut-based use of AI recommendations [19,21]. Students' orientation toward mastery (learning for competence) versus performance (learning for validation), alongside their prior technological experience, likely moderates the efficiency of these pathways [18,20,22].

In the context of rapidly expanding AI deployment in health systems, AI literacy is increasingly regarded as a prerequisite for safe and accountable digital health practice, with emerging evidence that targeted educational interventions can improve learners' understanding of AI concepts and their confidence in questioning algorithmic outputs [3,15]. Building on this work, the present study uses a three-wave longitudinal design across a 12-month clinical rotation to examine temporal associations among AI-assisted diagnosis participation, AI literacy, and medical critical thinking. Specifically, it investigates whether AI-assisted diagnosis participation is prospectively associated with subsequent critical thinking, whether AI literacy statistically mediates this association, and how these longitudinal pathways differ by students' prior technological experience and learning goal orientations. By doing so, the study aims to provide timely empirical guidance on how AI augmented curricula and clinical learning environments can be structured so that routine use of AI supports, rather than undermines, medical students' critical appraisal and diagnostic responsibility.

Critically, this study was conducted within a supervised clinical training context where experienced faculty mentors guided students' interactions with AI diagnostic tools. This supervisory structure is theorised to be essential for the hypothesised pathway from AI participation to literacy to critical thinking. Without close mentorship, learners—who are not yet experienced practitioners—may struggle to discern whether discrepancies between their own clinical judgment and AI-generated outputs reflect their own diagnostic errors or algorithmic limitations [6]. The presence of expert guidance creates a pedagogically "safe" learning environment where students can experiment with AI tools, receive immediate corrective feedback, and iteratively develop the critical evaluation skills that constitute AI literacy [32]. Thus, while we hypothesise that AI-assisted diagnosis participation fosters AI literacy, we emphasise that this

developmental process is unlikely to occur through passive or unsupervised exposure to AI systems. Throughout this manuscript, "AI-assisted diagnosis participation" refers specifically to engagement with AI tools under structured faculty supervision during clinical rotations.

Results

Preliminary Analyses and Measurement Validity

Analyses were based on the 372 students who completed all three measurement waves. Sample characteristics and attrition comparisons are summarised in Table 5. There were no statistically significant differences between the baseline and retained samples in age, gender, clinical year, cumulative GPA, prior technological experience, or learning goal orientations, and effect sizes for these comparisons were small (all $|d| \leq 0.05$; Table 5).

Descriptive statistics and bivariate correlations for the primary study variables are presented in Table 1. Mean AI-assisted diagnosis participation increased over time, from 3.12 (SD = 1.18) at baseline to 4.67 (SD = 1.09) at six months and 5.41 (SD = 0.98) at 12 months. AI literacy showed a similar pattern, with means of 3.45 (SD = 0.87), 4.89 (SD = 0.79), and 5.67 (SD = 0.73) at T_0 , T_1 , and T_2 , respectively. Medical critical thinking scores increased more modestly, from 4.23 (SD = 0.68) at baseline to 4.52 (SD = 0.65) at six months and 4.78 (SD = 0.64) at 12 months. Skewness values ranged from -0.45 to 0.38 and kurtosis values from -0.52 to 0.41 , suggesting that deviations from normality were limited and that the use of maximum likelihood estimation was appropriate.

Zero-order correlations showed substantial longitudinal stability within each construct (Table 1). For AI-assisted diagnosis participation, correlations across waves ranged from $r = .58$ to $.71$; for AI literacy, from $r = .54$ to $.73$; and for medical critical thinking, from $r = .61$ to $.74$. Correlations between AI-assisted diagnosis participation and AI literacy were moderate to large ($r = .38$ – $.74$), and correlations between AI-assisted diagnosis participation and critical thinking were smaller but consistently positive ($r = .25$ – $.42$). AI literacy was moderately associated with critical thinking at each wave ($r = .43$ – $.67$). Prior technological experience was positively related to AI participation, AI literacy, and critical thinking ($r = .27$ – $.43$), whereas performance orientation showed weak and partly negative correlations with critical thinking at later waves (e.g., $r = -.11$ at T_1 and $r = -.15$ at T_2 ; Table 1).

To examine whether AI-assisted diagnosis participation and AI literacy represented empirically distinct constructs, confirmatory factor analyses compared a two-factor solution (separating participation and literacy) with a more restrictive one-factor solution. The two-factor model showed clearly superior fit indices and lower information criteria than the one-factor model, supporting the conceptualisation of AI-assisted diagnosis participation as a behavioural construct and AI literacy as a competence construct.

Longitudinal measurement invariance testing supported configural, metric, and scalar invariance for AI-assisted diagnosis participation, AI literacy, and medical critical thinking across the three waves. These results indicate that the factor structures and response scales were stable over time and that comparisons of observed and latent means across waves are interpretable in terms of change rather than measurement artefacts.

Table 1 Bivariate Correlations Among Primary Variables (N = 372)

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. AI Participation T ₀	—											
2. AI Participation T ₁	.64**	—										
3. AI Participation T ₂	.58**	.71**	—									
4. AI Literacy T ₀	.49**	.41**	.38**	—								
5. AI Literacy T ₁	.52**	.68**	.61**	.59**	—							
6. AI Literacy T ₂	.47**	.63**	.74**	.54**	.73**	—						
7. Critical Thinking T ₀	.25**	.28**	.31**	.54**	.46**	.43**	—					
8. Critical Thinking T ₁	.27**	.35**	.38**	.48**	.58**	.55**	.68**	—				
9. Critical Thinking T ₂	.29**	.33**	.42**	.45**	.54**	.67**	.61**	.74**	—			
10. Prior Tech Experience	.41**	.35**	.32**	.43**	.38**	.35**	.31**	.29**	.27**	—		
11. Mastery Orientation	.33**	.36**	.39**	.37**	.41**	.44**	.46**	.48**	.51**	.24**	—	
12. Performance Orientation	.16**	.09	.05	.14**	.08	.03	-.06	-.11*	-.15**	.13*	-.21**	—
Descriptive Statistics												
M	3.12	4.67	5.41	3.45	4.89	5.67	4.23	4.52	4.78	4.31	5.78	4.12
SD	1.18	1.09	0.98	0.87	0.79	0.73	0.68	0.65	0.64	1.22	0.91	1.43

Note. N = 372 (complete cases analysis). **p < .01, *p < .05.

Longitudinal Relationships: Cross-Lagged Panel Analysis

A three-wave cross-lagged panel model (CLPM) was used to examine temporal associations among AI-assisted diagnosis participation, AI literacy, and medical critical thinking while controlling for baseline covariates (age, gender, and GPA). The model showed acceptable global fit to the data ($\chi^2 = 167.4$, $df = 89$, $RMSEA = 0.049$, $SRMR = 0.041$).

Autoregressive paths indicated high stability for all constructs across adjacent waves (Table 2). AI-assisted diagnosis participation at T₀ predicted participation at T₁ ($\beta = 0.640$, $SE = 0.039$, $p < .001$, 95% CI [0.564, 0.716]) and participation at T₁ predicted participation at T₂ ($\beta = 0.680$, $SE = 0.036$, $p < .001$, 95% CI [0.610, 0.750]). AI literacy showed comparable stability from T₀ to T₁

($\beta = 0.590$, $SE = 0.042$, $p < .001$, 95% CI [0.508, 0.672]) and from T_1 to T_2 ($\beta = 0.610$, $SE = 0.038$, $p < .001$, 95% CI [0.536, 0.684]). Medical critical thinking was also highly stable ($T_0 \rightarrow T_1$: $\beta = 0.710$, $SE = 0.033$, $p < .001$, 95% CI [0.646, 0.774]; $T_1 \rightarrow T_2$: $\beta = 0.740$, $SE = 0.031$, $p < .001$, 95% CI [0.680, 0.800]).

Cross-lagged paths are summarised in Table 2. Higher AI-assisted diagnosis participation at baseline was associated with higher AI literacy at six months ($\beta = 0.341$, $SE = 0.045$, $p < .001$, 95% CI [0.252, 0.428]) and higher AI literacy at six months was associated with higher critical thinking at 12 months ($\beta = 0.289$, $SE = 0.035$, $p < .001$, 95% CI [0.221, 0.359]). Baseline AI-assisted diagnosis participation also predicted higher critical thinking at six months ($\beta = 0.234$, $SE = 0.042$, $p < .001$, 95% CI [0.152, 0.316]). A similar pattern was observed for the second interval: AI-assisted diagnosis participation at T_1 predicted AI literacy at T_2 ($\beta = 0.290$, $SE = 0.041$, $p < .001$, 95% CI [0.210, 0.370]) and critical thinking at T_2 ($\beta = 0.218$, $SE = 0.039$, $p < .001$, 95% CI [0.142, 0.294]). In addition, AI literacy at T_0 predicted critical thinking at T_1 ($\beta = 0.160$, $SE = 0.046$, $p < .001$, 95% CI [0.070, 0.250]), indicating that higher initial literacy was associated with subsequent increases in critical thinking.

Although cross-lagged panel models cannot definitively establish causal direction [33,34], the longitudinal design permits examination of temporal precedence consistent with our theoretical framework. To assess whether the observed pattern of associations was equally consistent with an alternative temporal ordering, we specified a "reverse causation" model (i.e., Critical Thinking \rightarrow AI Literacy \rightarrow AI Participation). We conducted a sensitivity analysis to evaluate whether the observed longitudinal pattern was equally compatible with a reversed temporal ordering. Specifically, we estimated an alternative model in which baseline critical thinking predicted AI literacy at six months, which in turn predicted AI-assisted diagnosis participation at 12 months. Relative to the prespecified model, the alternative specification showed poorer fit based on the model comparison test reported in Table 2 (χ^2 difference = 31.7, $p < .001$) and a higher information criterion ($\Delta AIC = 31.7$). In addition, the key reversed cross-lagged paths were small and less consistent (critical thinking $T_0 \rightarrow$ AI literacy T_1 : $\beta = 0.089$, $SE = 0.043$, $p = .038$; AI literacy $T_1 \rightarrow$ AI participation T_2 : $\beta = 0.076$, $SE = 0.041$, $p = .065$). Taken together, these results indicate that the data are more consistent with the hypothesised ordering than with the fully reversed specification; however, unmeasured confounding and alternative causal processes cannot be ruled out in this observational design. Importantly, the interpretation of directional associations rests primarily on the theoretical framework derived from Situated Learning Theory and the time-ordered measurement design, rather than on model fit comparisons alone.

The total effect of AI participation on critical thinking development across the complete temporal sequence achieved substantial magnitude ($\beta = .237$, $SE = .041$, $p < .001$), indicating that a one-standard-deviation increase in AI participation corresponds to a .237 standard deviation enhancement in critical thinking competency. This effect size exceeds typical technology-learning

relationships documented in educational research while demonstrating practical significance within medical education contexts [23].

Table 2 Cross-Lagged Panel Analysis: Temporal Relationships (N = 372)

Pathway	β	SE	t	p	95% CI
Autoregressive Effects					
AI Participation T ₀ → AI Participation T ₁	.640	.039	16.41	<.001	[.564, .716]
AI Participation T ₁ → AI Participation T ₂	.680	.036	18.89	<.001	[.610, .750]
AI Literacy T ₀ → AI Literacy T ₁	.590	.042	14.05	<.001	[.508, .672]
AI Literacy T ₁ → AI Literacy T ₂	.610	.038	16.05	<.001	[.536, .684]
Critical Thinking T ₀ → Critical Thinking T ₁	.710	.033	21.52	<.001	[.646, .774]
Critical Thinking T ₁ → Critical Thinking T ₂	.740	.031	23.87	<.001	[.680, .800]
Cross-Lagged Effects (Theoretical Model)					
AI Participation T ₀ → AI Literacy T ₁	.341	.045	7.56	<.001	[.252, .428]
AI Participation T ₁ → AI Literacy T ₂	.290	.041	7.07	<.001	[.210, .370]
AI Literacy T ₀ → Critical Thinking T ₁	.160	.046	3.48	<.001	[.070, .250]
AI Literacy T ₁ → Critical Thinking T ₂	.289	.035	8.29	<.001	[.221, .359]
AI Participation T ₀ → Critical Thinking T ₁	.234	.042	5.57	<.001	[.152, .316]
AI Participation T ₁ → Critical Thinking T ₂	.218	.039	5.59	<.001	[.142, .294]
AI Participation T ₀ → Critical Thinking T ₂ (Total)	.237	.041	5.78	<.001	[.157, .317]
Cross-Lagged Effects (Alternative Model)					
Critical Thinking T ₀ → AI Literacy T ₁	.089	.044	2.02	.044	[.003, .175]
Critical Thinking T ₁ → AI Literacy T ₂	.072	.041	1.76	.079	[-.008, .152]
AI Literacy T ₀ → AI Participation T ₁	.127	.047	2.70	.007	[.035, .219]
AI Literacy T ₁ → AI Participation T ₂	.105	.042	2.50	.012	[.023, .187]
Critical Thinking T ₀ → AI Participation T ₁	.089	.043	2.07	.038	[.005, .173]
Critical Thinking T ₁ → AI Participation T ₂	.076	.041	1.85	.065	[-.005, .157]
Within-Time Correlations					
AI Participation ↔ AI Literacy (T ₀)	.490	.041	11.95	<.001	[.410, .570]
AI Participation ↔ Critical Thinking (T ₀)	.250	.048	5.21	<.001	[.156, .344]
AI Literacy ↔ Critical Thinking (T ₀)	.540	.038	14.21	<.001	[.466, .614]
AI Participation ↔ AI Literacy (T ₁)	.680	.034	20.00	<.001	[.614, .746]
AI Participation ↔ Critical Thinking (T ₁)	.350	.045	7.78	<.001	[.262, .438]
AI Literacy ↔ Critical Thinking (T ₁)	.580	.037	15.68	<.001	[.508, .652]

Pathway	β	SE	t	p	95% CI
AI Participation \leftrightarrow AI Literacy (T ₂)	.740	.032	23.13	<.001	[.678, .802]
AI Participation \leftrightarrow Critical Thinking (T ₂)	.420	.043	9.77	<.001	[.336, .504]
AI Literacy \leftrightarrow Critical Thinking (T ₂)	.670	.035	19.14	<.001	[.602, .738]
Model Fit Indices					
Theoretical Model					
χ^2 (df)	167.4(89)	—	—	<.001	—
CFI	.951	—	—	—	—
TLI	.943	—	—	—	—
RMSEA [90% CI]	.049 [.037, .061]	—	—	—	—
SRMR	.041	—	—	—	—
Alternative Model					
χ^2 (df)	199.1(89)	—	—	<.001	—
CFI	.934	—	—	—	—
TLI	.921	—	—	—	—
RMSEA [90% CI]	.058 [.046, .070]	—	—	—	—
SRMR	.056	—	—	—	—
Model Comparison					
$\Delta\chi^2$ (df)	31.7(0)	—	—	<.001	—
Δ AIC	31.7	—	—	—	—

Longitudinal Mediation Analysis

To examine whether the longitudinal association between AI-assisted diagnosis participation and medical critical thinking was statistically mediated by AI literacy, we specified a three-wave mediation model focusing on the T₀ \rightarrow T₁ \rightarrow T₂ temporal sequence (Table 3). Specifically, this model tested whether baseline AI-assisted diagnosis participation was associated with AI literacy at six months, and whether AI literacy at six months was in turn associated with critical thinking at 12 months, after controlling for prior levels of all constructs. In this model, baseline AI-assisted diagnosis participation retained a modest direct association with critical thinking at 12 months after inclusion of the mediator pathway (direct $\beta = 0.095$, SE = 0.039, $p = .015$, 95% CI [0.018, 0.172]). Baseline AI-assisted diagnosis participation was positively related to AI literacy at six months ($\beta = 0.341$, SE = 0.038, $p < .001$, 95% CI [0.267, 0.415]), and AI literacy at six months was positively related to critical thinking at 12 months ($\beta = 0.289$, SE = 0.035, $p < .001$, 95% CI [0.220, 0.358]).

The indirect pathway was statistically significant: baseline AI-assisted diagnosis participation was associated with higher AI literacy at six months ($\beta = 0.341$, $SE = 0.038$, $p < .001$), and AI literacy at six months was associated with higher critical thinking at 12 months ($\beta = 0.289$, $SE = 0.035$, $p < .001$), yielding an indirect effect of $\beta = 0.142$ ($SE = 0.019$, $p < .001$, 95% CI [0.089, 0.201]). The total association between baseline AI-assisted diagnosis participation and critical thinking at 12 months was $\beta = 0.237$ ($SE = 0.041$, $p < .001$, 95% CI [0.156, 0.318]). Approximately 38% of this total association (95% CI [0.26, 0.54]) was transmitted through the sequential pathway involving AI literacy, indicating partial mediation.

Table 3 Mediation Analysis: AI Literacy as Mediator (N = 372)

Effect Type and Pathway	Estimate	SE	p	95% CI
Direct Effects				
AI Participation T ₀ → Critical Thinking T ₂	.095	.039	.015	[.018, .172]
AI Participation T ₀ → AI Literacy T ₁	.341	.038	<.001	[.267, .415]
AI Literacy T ₁ → Critical Thinking T ₂	.289	.035	<.001	[.220, .358]
Indirect Effects (Monte Carlo CI)				
AI Part. T ₀ → AI Lit. T ₁ → Crit. Think. T ₂	.142	.019	<.001	[.089, .201]
Total Effects				
AI Participation T ₀ → Critical Thinking T ₂	.237	.041	<.001	[.156, .318]
Mediation Assessment				
Proportion Mediated	.38	.07	—	[.26, .54]
Sobel Test	$z = 5.74$, $p < .001$	—	—	—
Alternative Bootstrap Methods				
Bias-Corrected Bootstrap (10,000 samples)	—	—	—	[.091, .203]
Percentile Bootstrap (10,000 samples)	—	—	—	[.087, .198]
Model Comparisons				
Partial Mediation Model				
χ^2 (df)	145.3(88)	—	<.001	—
AIC	15623.7	—	—	—
CFI	.951	—	—	—
RMSEA	.043	—	—	—
Complete Mediation Model				
χ^2 (df)	154.0(89)	—	<.001	—
AIC	15630.4	—	—	—
CFI	.947	—	—	—

Effect Type and Pathway	Estimate	SE	p	95% CI
RMSEA	.045	—	—	—
No Mediation Model				
χ^2 (df)	178.6(90)	—	<.001	—
AIC	15652.9	—	—	—
CFI	.938	—	—	—
RMSEA	.051	—	—	—
Model Comparison Tests				
Partial vs. Complete Mediation	$\Delta\chi^2(1) = 8.7, p = .003$	—	—	—
Partial vs. No Mediation	$\Delta\chi^2(2) = 33.3, p < .001$	—	—	—

This mediation proportion represents a moderate-to-large effect within educational technology research contexts, where indirect effects typically account for 20-35% of total associations [24]. Approximately 38% of the total longitudinal association between baseline AI-assisted diagnosis participation and critical thinking at 12 months (95% CI [0.26, 0.54]) was transmitted through AI literacy at six months, indicating partial mediation. This proportion suggests that AI literacy accounts for a meaningful share of the prospective association, while a statistically significant direct path remained ($\beta = .095, SE = .039, p = .015$), implying that additional pathways may also link participation with later critical thinking. The magnitude suggests that the sequential pathway from AI participation through AI literacy to critical thinking constitutes a substantial portion of the overall longitudinal association. That is, students who participated more actively in AI-assisted diagnosis tended to show higher AI literacy at the subsequent time point, and this higher literacy was in turn associated with higher critical thinking at the following time point. Because mediation estimates in observational longitudinal data remain vulnerable to unmeasured confounding, these proportions should be interpreted as describing statistical decomposition of associations rather than definitive mechanistic attribution. The remaining direct association indicates that participation may also relate to critical thinking through mechanisms not captured by AI literacy, warranting further investigation.

Monte Carlo confidence interval procedures with 10,000 resamples confirmed mediation effect stability across alternative estimation methods. Bias-corrected accelerated intervals (95% CI [.091, .203]) and percentile intervals (95% CI [.087, .198]) yielded similar ranges, supporting confidence in indirect effect estimates. Sensitivity analysis through varying distributional assumptions and missing data treatments produced consistent results, indicating robustness across analytical choices.

The partial mediation model demonstrated superior fit compared to complete mediation alternatives ($\Delta\chi^2 = 8.7, df = 1, p = .003$), indicating meaningful direct effects alongside indirect pathways. The direct effect of AI participation on critical thinking remained significant ($\beta = .095,$

SE = .039, $p = .015$) after accounting for AI literacy mediation, suggesting that additional mechanisms beyond AI literacy contribute to AI participation effectiveness for critical thinking development.

Moderation by Individual Differences

Multi-group structural equation models were used to examine whether the longitudinal associations via AI literacy differed by prior technological experience and learning goal orientation (Table 4). The model revealed a "Matthew Effect." For prior technological experience, the sample was split at the median into lower-experience ($n = 186$) and higher-experience ($n = 186$) groups. In the lower-experience group, the path from AI-assisted diagnosis participation to AI literacy was $\beta = 0.287$ (SE = 0.049, $p < .001$), the path from AI literacy to critical thinking was $\beta = 0.271$ (SE = 0.054, $p < .001$), and the indirect effect via AI literacy was $\beta = 0.096$ (SE = 0.029, $p < .001$, 95% CI [0.048, 0.155]). In the higher-experience group, the corresponding paths were somewhat larger: AI-assisted diagnosis participation to AI literacy $\beta = 0.389$ (SE = 0.045, $p < .001$), AI literacy to critical thinking $\beta = 0.324$ (SE = 0.048, $p < .001$), and the indirect effect $\beta = 0.178$ (SE = 0.031, $p < .001$, 95% CI [0.121, 0.245]). The direct path from AI-assisted diagnosis participation to critical thinking was small in both groups ($\beta = 0.108$, SE = 0.051, $p = .034$ in the low-experience group; $\beta = 0.089$, SE = 0.047, $p = .060$ in the high-experience group). A χ^2 difference test comparing a constrained model ($\chi^2(156) = 198.7$, $p = .008$) with an unconstrained model ($\chi^2(152) = 186.3$, $p = .024$) indicated that allowing the key structural paths to vary by technological experience improved model fit ($\Delta\chi^2(4) = 12.4$, $p = .015$). Supplementary continuous-variable moderation analyses using latent moderated structural equations corroborated these patterns: the interaction between AI participation and technological experience significantly predicted AI literacy, indicating that the association between participation and literacy was stronger at higher levels of technological experience (see Supplementary Table S3).

For learning goal orientation, students were divided into lower- and higher-mastery groups using a median split on mastery orientation scores ($n = 185$ and $n = 187$, respectively). In the lower-mastery group, AI-assisted diagnosis participation was associated with AI literacy ($\beta = 0.298$, SE = 0.051, $p < .001$) and AI literacy was associated with critical thinking ($\beta = 0.247$, SE = 0.058, $p < .001$), yielding an indirect effect of $\beta = 0.081$ (SE = 0.023, $p < .001$, 95% CI [0.032, 0.139]). The direct path from AI-assisted diagnosis participation to critical thinking was $\beta = 0.134$ (SE = 0.053, $p = .012$). In the higher-mastery group, the path from AI-assisted diagnosis participation to AI literacy was $\beta = 0.412$ (SE = 0.043, $p < .001$), the path from AI literacy to critical thinking was $\beta = 0.354$ (SE = 0.046, $p < .001$), and the indirect effect via AI literacy was $\beta = 0.189$ (SE = 0.027, $p < .001$, 95% CI [0.134, 0.256]). The direct path from AI-assisted diagnosis participation to critical thinking was smaller and not statistically significant in the higher-mastery group ($\beta = 0.067$, SE = 0.045, $p = .137$). A multi-group comparison showed that a model allowing these paths to differ by mastery orientation ($\chi^2(152) = 187.7$, $p = .021$) fit better than a constrained model ($\chi^2(156) = 203.4$, $p = .005$; $\Delta\chi^2(4) = 15.7$, $p = .003$). Continuous-variable moderation analyses yielded

convergent findings: the AI Literacy \times Mastery Orientation interaction significantly predicted critical thinking, confirming that higher mastery orientation strengthened the association between AI literacy and critical thinking outcomes. Full results from simple slope analyses and conditional indirect effects are presented in Supplementary Table S3.

To examine combined moderation, students were initially classified into four groups defined by the intersection of prior technological experience (low vs high) and mastery orientation (low vs high) using median splits. The estimated indirect effect from AI-assisted diagnosis participation to critical thinking via AI literacy varied across these groups (Table 4). In the low-experience/low-mastery group, the indirect effect was $\beta = 0.052$ (SE = 0.019, $p = .005$, 95% CI [0.018, 0.095]); in the low-experience/high-mastery group, $\beta = 0.139$ (SE = 0.026, $p < .001$, 95% CI [0.091, 0.189]); in the high-experience/low-mastery group, $\beta = 0.125$ (SE = 0.024, $p < .001$, 95% CI [0.084, 0.174]); and in the high-experience/high-mastery group, $\beta = 0.234$ (SE = 0.037, $p < .001$, 95% CI [0.168, 0.312]). A model allowing these indirect paths to vary across the four groups ($\chi^2(152) = 201.3$, $p = .003$) fit better than a model constraining them to be equal ($\chi^2(164) = 234.8$, $p < .001$; $\Delta\chi^2(12) = 33.5$, $p = .001$). We also estimated a continuous interaction model including both moderators simultaneously (AI Participation \times Technological Experience and AI Literacy \times Mastery Orientation). The three-way conditional indirect effect at high levels of both moderators (+1 SD) was $\beta = 0.221$ (SE = 0.034, $p < .001$, 95% CI [0.158, 0.291]), closely approximating the median-split estimate for the high-experience/high-mastery group ($\beta = 0.234$). Conversely, the conditional indirect effect at low levels of both moderators (-1 SD) was $\beta = 0.067$ (SE = 0.022, $p = .002$, 95% CI [0.027, 0.112]), consistent with the median-split low-experience/low-mastery estimate ($\beta = 0.052$). These convergent findings across analytical approaches support the robustness of the moderation patterns and suggest that the median-split results provide a reasonable approximation of the continuous moderation effects.

Moderation analysis revealed that the magnitude of the indirect pathway varied across learner profiles. For high-experience, mastery-oriented students, the sequential association from participation through literacy to critical thinking was strongest ($\beta = .234$, 95% CI [.168, .312]): in this subgroup, greater AI participation was most strongly associated with subsequent AI literacy gains, and AI literacy was most strongly associated with subsequent critical thinking gains. Conversely, students with limited technological experience and performance orientation showed weaker indirect associations ($\beta = .052$, 95% CI [.018, .095]), suggesting that the pathway from participation through literacy to critical thinking may be attenuated for learners who approach AI tools primarily as efficiency aids rather than learning opportunities. Notably, the direct association between AI Participation and Critical Thinking was non-significant for the performance-oriented group ($p > .05$), indicating that for these students, participation was primarily associated with critical thinking through its association with AI literacy development, rather than through direct mechanisms. These patterns highlight the importance of considering individual differences in designing AI-integrated curricula to promote equitable learning outcomes.

To address potential limitations associated with the categorization of continuous variables, we conducted a robustness check using Latent Moderated Structural Equations (LMS)[35,36]. This analysis estimated the interaction effects of Prior Technological Experience and Mastery Orientation on the structural paths without artificial grouping. Specifically, we estimated models including latent interaction terms: (a) AI Participation \times Prior Technological Experience predicting AI Literacy, and (b) AI Literacy \times Mastery Orientation predicting Critical Thinking. The LMS results corroborated the multi-group findings: the interaction term between AI Participation and Prior Technological Experience significantly predicted AI Literacy, and the interaction between AI Literacy and Mastery Orientation was similarly significant. Simple slope analyses confirmed that the association between participation and literacy strengthened as experience and mastery orientation increased (Supplementary Table S3)[37]. These continuous variable analyses confirm that the observed "Matthew Effect" is not an artifact of discretization.

Table 4 Individual Difference Moderation Analysis (N = 372)

Moderator Configuration	Low Experience (n = 186)	High Experience (n = 186)	Low Mastery (n = 185)	High Mastery (n = 187)
Prior Technological Experience Moderation				
Direct Effects				
AI Participation \rightarrow AI Literacy	.287(.049)**	.389(.045)**	—	—
AI Literacy \rightarrow Critical Thinking	.271(.054)**	.324(.048)**	—	—
AI Participation \rightarrow Critical Thinking	.108(.051)*	.089(.047)	—	—
Indirect Effect				
Mediation Pathway	.096(.029)**	.178(.031)**	—	—
95% Confidence Interval	[.048, .155]	[.121, .245]	—	—
Learning Goal Orientation Moderation				
Direct Effects				
AI Participation \rightarrow AI Literacy	—	—	.298(.051)**	.412(.043)**
AI Literacy \rightarrow Critical Thinking	—	—	.247(.058)**	.354(.046)**
AI Participation \rightarrow Critical Thinking	—	—	.134(.053)*	.067(.045)
Indirect Effect				
Mediation Pathway	—	—	.081(.023)**	.189(.027)**
95% Confidence Interval	—	—	[.032, .139]	[.134, .256]
Combined Moderation Analysis				
Low Experience \times Low Mastery	.052(.019)**	95% CI [.018, .095]		
Low Experience \times High Mastery	.139(.026)**	95% CI [.091, .189]		
High Experience \times Low Mastery	.125(.024)**	95% CI [.084, .174]		

Moderator Configuration	Low Experience (n = 186)	High Experience (n = 186)	Low Mastery (n = 185)	High Mastery (n = 187)
High Experience × High Mastery	.234(.037)**	95% CI [.168, .312]		
Multi-Group Comparisons				
Prior Experience				
Constrained Model	$\chi^2(156) = 198.7, p = .008$			
Unconstrained Model	$\chi^2(152) = 186.3, p = .024$			
Difference Test	$\Delta\chi^2(4) = 12.4, p = .015$			
Goal Orientation				
Constrained Model	$\chi^2(156) = 203.4, p = .005$			
Unconstrained Model	$\chi^2(152) = 187.7, p = .021$			
Difference Test	$\Delta\chi^2(4) = 15.7, p = .003$			
Combined Moderation				
Constrained Model	$\chi^2(164) = 234.8, p < .001$			
Unconstrained Model	$\chi^2(152) = 201.3, p = .003$			
Difference Test	$\Delta\chi^2(12) = 33.5, p = .001$			

Note. Unstandardized coefficients with standard errors in parentheses. ** $p < .01$, * $p < .05$. Groups divided by median split procedures for Prior Technological Experience ($Mdn = 4.25$) and Learning Goal Orientation composite scores (Mastery - Performance orientation difference scores, $Mdn = 1.42$). Supplementary continuous-variable moderation analyses using latent moderated structural equations (LMS) are reported in Supplementary Table S3; these analyses yielded convergent findings.

Model Comparison and Validation

Additional model comparison and sensitivity analyses are summarised in the Supplementary Material (Supplementary Table S2). We compared the proposed partial mediation model (M2; $\chi^2 = 145.3$, $df = 88$, $AIC = 15623.7$, $CFI = .951$, $RMSEA = .043$) with a set of theoretically motivated alternatives, including no mediation (M1), complete mediation (M3), reverse causation (M4), bidirectional cross-lagged effects (M5), concurrent mediation (M6), and a multiple-mediator specification (M7). Models that removed the AI literacy pathway or imposed a reverse temporal ordering showed poorer fit and higher information criteria (e.g., M1 $AIC = 15652.9$; M4 $AIC = 15667.8$; M6 $AIC = 15643.6$). More complex extensions that added reciprocal paths or additional mediators yielded very similar global fit (M5 $AIC = 15625.2$; M7 $AIC = 15626.1$) and did not provide a statistically meaningful improvement over M2 based on the chi-square difference tests reported in Supplementary Table S2 (M5: $\Delta\chi^2 = 2.5$, $p = .287$; M7: $\Delta\chi^2 = 5.8$, $p = .214$). We therefore retain M2 as the primary specification on grounds of parsimony and a priori theoretical motivation. While model fit indices alone cannot definitively establish causal agency [33], the

temporal precedence established by the longitudinal design supports the theoretical plausibility of the hypothesized directional paths over reverse-causation alternatives. Accordingly, directional interpretations in this study are grounded primarily in the theoretical framework (Situated Learning Theory), the temporal ordering of measurements, and the a priori specification of hypotheses, rather than in post-hoc model comparisons alone. The cross-lagged panel design controls for prior levels of each construct and permits examination of whether associations are consistent with theoretical expectations, but cannot rule out unmeasured confounding or alternative causal processes.

Sensitivity analyses evaluated the robustness of the findings to alternative treatments of missing data, outliers, and model parameterisation. Analyses based on complete cases, analyses excluding univariate and multivariate outliers, and models with alternative sets of auxiliary variables yielded parameter estimates and confidence intervals that were highly similar to those of the primary FIML models. Finally, cross-validation procedures using random 60/40 splits of the sample showed that key standardized path coefficients and indirect effects were of comparable magnitude and statistical significance across subsamples, supporting the stability of the longitudinal patterns observed in the full dataset (Supplementary Table S2).

Discussion

In this three-wave longitudinal study, higher levels of AI-assisted diagnosis participation were prospectively associated with higher medical critical thinking scores over a 12-month clinical rotation, after accounting for prior levels of all constructs and baseline covariates. These patterns do not support a simple “cognitive atrophy” view in which greater exposure to automated systems necessarily coincides with declines in human reasoning [7-8]. Instead, within a supervised clinical training context, students who were more actively involved in AI-assisted diagnosis tended to show stronger subsequent critical thinking. At the same time, these findings should not be interpreted as evidence that any particular AI system improves diagnostic reasoning. The design remains observational and relies on self-reported measures, and unmeasured factors such as general cognitive ability or conscientiousness may influence both AI participation and critical thinking. Rather than evaluating the effectiveness of a specific AI tool, the results speak to how AI augmented learning environments can be aligned with the development of analytic and reflective diagnostic practices when students are encouraged to engage critically with algorithmic recommendations [5,6,14].

A central contribution of this analysis is the identification of AI literacy as a statistically mediating variable in the longitudinal association between students’ behavioural engagement with AI-assisted diagnosis (“participation”) and later critical thinking, thereby distinguishing technology usage from digital competence. Specifically, our model disentangles “participation” (behavioural frequency of interaction with AI-assisted diagnostic tasks) from “AI literacy” (the acquired competence to comprehend, evaluate, and appropriately integrate algorithmic outputs), addressing a common limitation in prior work that conflates use with competence. Drawing on Situated

Learning Theory, the observed sequential pattern of associations is consistent with a developmental trajectory in which repeated, supervised participation (“learning by doing”) provides situated opportunities for students to actively acquire domain-relevant AI literacy [11], which in turn is associated with more reflective and analytic reasoning and may serve as an intermediary competence linking participation behaviour to reasoning outcomes. Importantly, because these inferences are based on observational longitudinal mediation models, the indirect pathway should be interpreted as a decomposition of associations rather than definitive evidence of a causal mechanism; the results are consistent with, but do not prove, the theorised account. Future research using experimental or quasi-experimental variation in AI-supported training, objective performance measures, and richer confounder control is needed to test whether strengthening AI literacy causally amplifies the benefits of AI participation for clinical reasoning.

Students who understand the probabilistic nature of AI outputs are better equipped to interrogate diagnostic suggestions, identify algorithmic hallucinations, and synthesize discordant data [15]. Rather than treating AI literacy as a fixed prerequisite for using AI systems, it was conceptualised as a developing competence that can change during training. The cross lagged models indicated that higher participation was followed by higher AI literacy, and higher AI literacy was in turn associated with later gains in critical thinking, yielding a statistically significant indirect pathway across the three waves. This pattern is consistent with the idea that frequent, active participation in AI-assisted diagnosis can provide opportunities for students to learn how AI systems are built and evaluated, how to interpret uncertainty in algorithmic outputs, and how to integrate AI recommendations into their own diagnostic reasoning [5,11]. This finding clarifies that AI literacy is not merely a moderator that amplifies the effects of usage, but a developmental outcome of usage itself that subsequently enables higher-order reasoning. While theoretically plausible as a moderator, our explicit testing of temporal precedence via cross-lagged modeling confirms that AI literacy emerges from participation, structurally functioning as a mediator rather than a pre-existing boundary condition. The mediation analyses should nonetheless be viewed as evidence of an indirect longitudinal association rather than a definitive psychological mechanism. Future studies that combine performance-based AI literacy tasks, objective usage logs, and experimental or quasi-experimental designs will be needed to test whether targeted AI literacy interventions can causally strengthen the link between AI participation and critical thinking development.

The analysis of individual differences reveals that this developmental pathway is constrained by prior technological experience, reflecting a “Matthew Effect” in digital medical education where the “rich get richer.” Students with high prior experience derived significantly greater critical thinking benefits from AI participation compared to their less experienced peers. This aligns with Cognitive Load Theory, suggesting that novices may expend excessive cognitive resources merely navigating the interface and interpreting basic algorithmic outputs, leaving limited capacity for the metacognitive evaluation required for critical thinking [25]. Consequently, medical curricula cannot adopt a “one-size-fits-all” approach to AI integration. For students with limited

technological backgrounds, foundational training in algorithmic logic must precede intensive clinical AI immersion to prevent cognitive overload and ensure equitable learning outcomes.

Furthermore, the moderating role of learning goal orientation—interpreted through Achievement Goal Theory—highlights the importance of student motivation in technology-mediated learning [17,18,19]. Mastery-oriented students, who view AI interactions as opportunities to understand the "why" behind a diagnosis, demonstrated the strongest mediation effects. In contrast, performance-oriented students, who may prioritize obtaining the "correct" diagnosis to validate competence, showed weaker critical thinking gains. This suggests that performance-oriented students may be more susceptible to "automation bias," accepting AI outputs at face value to achieve rapid results [18-19]. To mitigate this, medical educators should restructure assessment strategies. Rather than evaluating students solely on diagnostic accuracy, assessments should require "explainable AI" tasks where students must explicitly justify their acceptance or rejection of an AI suggestion [20]. This pedagogical shift forces performance-oriented learners to engage with the reasoning process, potentially shifting their focus from outcome validation to competency development [9]. Although the statistical models treat AI literacy as a mediator rather than a moderator, the conditional pattern of associations should be interpreted as highlighting when and for whom AI participation is most likely to be linked with gains in critical thinking, rather than as evidence of a separate interaction effect.

Contextualizing these findings within the specific educational setting of the study enhances their transferability to global medical training. The participants were 4th and 5th-year students in the Chinese medical education system, engaging in full-time clinical rotations comparable to the "sub-internship" or early residency phases in North American and European models [26]. In this high-volume clinical environment, students frequently encounter diverse pathologies and are tasked with preliminary diagnostic workups under supervision. This intense "immersion" provides a unique ecological niche for AI integration, as the high case volume necessitates efficient tool use while offering ample iterations for "trial-and-error" learning with AI systems. Therefore, while the academic titles differ, the functional role of these students mirrors that of junior residents in other systems, suggesting that the observed benefits of AI participation on critical thinking are likely generalizable to postgraduate medical training contexts globally where clinical responsibility and AI tool availability intersect.

Based on these insights, we propose a multi-tiered framework for implementing AI in medical curricula, with supervision as the foundational prerequisite. First, stratified training is essential: novices require explicit instruction on AI limitations and prompt engineering before clinical application to reduce cognitive load, while advanced students should engage in "AI audit" exercises—deliberately identifying errors in AI diagnostic reports under faculty observation. Second, institutions must move beyond technical training to foster "epistemic agency," teaching students to treat AI as a second opinion requiring validation rather than a gold standard for clinical judgment [2]. Third, and most critically, faculty development is paramount; clinical preceptors

must be trained to model "skeptical usage" behaviours, explicitly verbalising their own critical evaluation of AI outputs during ward rounds and providing real-time feedback when students accept AI recommendations uncritically. This cultural shift, supported by institutional policy, ensures that AI remains a servant to clinical reasoning rather than its substitute. We emphasise that the observed benefits of AI participation on critical thinking likely depend on this supervisory infrastructure; implementing AI tools without concurrent investment in faculty training and structured oversight may fail to produce similar outcomes and could potentially undermine reasoning development.

Several limitations should be considered when interpreting these results. First, despite the longitudinal design and the use of cross-lagged panel models, the data remain observational and rely on self-reported measures of AI-assisted diagnosis participation, AI literacy, and medical critical thinking. As such, the analyses cannot establish causal effects of AI tools or AI literacy, and common method variance may inflate some of the associations observed. Relatedly, while the longitudinal design and cross-lagged panel models provide evidence of temporal precedence and statistical mediation, these analyses remain fundamentally correlational and cannot definitively establish that AI participation causally increases AI literacy, which in turn causally enhances critical thinking [33,34]. The interpretation of AI literacy as a mediating mechanism is grounded in theoretical reasoning from Situated Learning Theory and the temporal structure of the three-wave design, rather than in experimental manipulation or model fit comparisons that cannot adjudicate causal direction. Future research employing experimental or quasi-experimental designs—such as randomised controlled trials of AI literacy interventions—would provide stronger evidence for causal claims. This concern was partially addressed by modelling distinct constructs over three time points and testing longitudinal measurement invariance, but future research should incorporate objective indicators such as system usage logs, performance-based AI literacy tasks, and standardized assessments of diagnostic reasoning to triangulate the findings [10,27,34]. Second, the study was conducted in three universities within one national and cultural context using a specific AI-assisted diagnosis system embedded in high-volume teaching hospitals. Patterns of AI adoption, supervision norms, and digital infrastructure may differ in other countries, specialties, or institutional settings, which could influence both AI usage and educational outcomes [14]. Third, although key demographic and academic variables were controlled and several individual differences examined, other unmeasured factors—such as general cognitive ability, baseline clinical reasoning skills, faculty mentorship quality, and local curriculum design—may also shape how AI participation translates into AI literacy and critical thinking. Taken together, these limitations underscore the need for multi-site, multi-method studies that combine longitudinal designs with richer behavioural and performance data to better understand how AI-augmented learning environments influence the development of critical thinking in medical education.

Overall, this study contributes to ongoing debates in digital medicine and medical education about whether AI integration will erode or enhance key professional capacities. Across one year of

supervised clinical training, more active participation in AI-assisted diagnosis was associated with gains in medical students' critical thinking, partly through the pathway involving AI literacy development, and these associations were stronger among students with higher technological experience and mastery-oriented goals. These findings support viewing AI systems as learning resources that can scaffold, rather than replace, human diagnostic reasoning—but only when deployed within thoughtfully designed educational contexts that include structured faculty supervision, explicit AI literacy curricula, and assessment strategies that reward critical evaluation of algorithmic outputs. The benefits observed in this study should not be assumed to generalise to unsupervised settings. At the same time, they highlight that benefits are not automatic and can be unevenly distributed, reinforcing the importance of explicit AI literacy training, careful attention to individual differences, and institutional safeguards that keep clinical responsibility and critical judgment firmly in human hands [5,6,14,27].

Methods

Study Design and Educational Context

This investigation employed a prospective longitudinal cohort design over twelve months with three measurement waves: baseline (T_0), mid-point (T_1 , 6 months), and follow-up (T_2 , 12 months). The timing of the assessments was chosen to capture changes in medical critical thinking, which typically require extended periods of training to manifest measurable differences [25]. A one-month AI system familiarization period preceded baseline data collection so that students had sufficient exposure to the AI-assisted diagnosis tools before initial assessment.

The study was conducted across three medical universities in China. Participants were fourth- and fifth-year medical students engaged in full-time clinical rotations at large teaching hospitals. In the Chinese medical education system, these years involve intensive clinical immersion in which students perform preliminary patient workups, contribute to ward rounds, and work under close supervision. Functionally, this role is comparable to sub-interns or early postgraduate trainees in many international systems. During rotations, students had supervised access to hospital information systems and routinely used AI tools integrated into the clinical workflow [26]. Specifically, they engaged with (1) computer-aided diagnosis systems for radiological image interpretation (for example, detection of pulmonary nodules on CT scans) and (2) AI-driven clinical decision support systems embedded in the electronic health record that generate differential diagnosis suggestions based on structured symptom and examination data. This high-volume clinical environment ensured that AI-assisted diagnosis participation reflected authentic, practice-embedded interaction with diagnostic algorithms rather than simulated classroom exercises.

Participants

A stratified cluster random sampling method was used to recruit participants, balancing representation across internal medicine, surgery, and comprehensive care rotations. Comprehensive power analysis employing G*Power 3.1.9 indicated $N \geq 420$ required for detecting theoretically meaningful effect sizes ($\beta \geq .25$) in longitudinal mediation models with 80% statistical power, $\alpha = .05$, and anticipated attrition $\leq 20\%$ across twelve-month intervals. The final analytical sample comprised 450 participants at baseline with 82.7% retention ($n = 372$) through final assessment, exceeding minimum requirements for stable parameter estimation in complex longitudinal models.

Attrition analysis indicated no significant differences in demographic characteristics or baseline levels of AI-assisted diagnosis participation, AI literacy, or medical critical thinking between retained and lost participants. Little's MCAR test ($\chi^2 = 45.67$, $p = .32$) supported the assumption that missingness was consistent with a Missing at Random mechanism. The mean age at baseline was 22.47 years ($SD = 1.73$), and 58.9% of the sample were women, reflecting the gender distribution of the regional medical student population. Comprehensive demographic characteristics and attrition statistics are reported in Table 5. Stratification considered institutional type (research-intensive vs teaching-focused), geographic distribution, and clinical specialization tracks to enhance external validity while maintaining representative sampling of the target population.

Table 5 Sample Characteristics and Demographic Profile (N = 450)

Characteristic	Baseline Sample (T ₀) n = 450	Retained Sample (T ₂) n = 372 (82.7%)	Attrition Analysis Statistical Comparison
Demographic Variables			
Age (years)			
Mean \pm SD	22.47 \pm 1.73	22.39 \pm 1.69	$t(448) = 0.84$, $p = .401$
Range	20-26	20-26	$d = 0.05$
Gender, n (%)			
Female	263 (58.4)	219 (58.9)	$\chi^2(1) = 0.11$, $p = .740$
Male	187 (41.6)	153 (41.1)	$\phi = 0.02$
Clinical Year, n (%)			
Fourth Year	201 (44.7)	167 (44.9)	$\chi^2(1) = 0.02$, $p = .888$
Fifth Year	249 (55.3)	205 (55.1)	$\phi = 0.01$
Academic Performance Indicators			
Cumulative GPA			
Mean \pm SD	3.52 \pm 0.39	3.54 \pm 0.38	$t(448) = -0.93$, $p = .354$
Range	2.67-4.00	2.71-4.00	$d = 0.05$

Characteristic	Baseline Sample (T ₀)	Retained Sample (T ₂)	Attrition Analysis
Clinical Rotation Experience (months)			
Mean ± SD	9.34 ± 2.41	9.45 ± 2.38	t(448) = -0.78, p = .436
Range	5-16	5-16	d = 0.05
Institutional Characteristics			
Institution Type, n (%)			
Research-Intensive	287 (63.8)	239 (64.2)	χ ² (1) = 0.09, p = .764
Teaching-Focused	163 (36.2)	133 (35.8)	φ = 0.01
Geographic Distribution, n (%)			
Urban	298 (66.2)	248 (66.7)	χ ² (1) = 0.12, p = .729
Peri-urban	152 (33.8)	124 (33.3)	φ = 0.02
Specialization Track, n (%)			
Internal Medicine	167 (37.1)	139 (37.4)	χ ² (2) = 0.23, p = .891
Surgery	152 (33.8)	125 (33.6)	Cramer's V = 0.02
Comprehensive Care	131 (29.1)	108 (29.0)	
Baseline Technology Experience			
Prior AI System Exposure, n (%)			
None	167 (37.1)	134 (36.0)	χ ² (2) = 1.12, p = .571
Limited (1-6 months)	198 (44.0)	168 (45.2)	Cramer's V = 0.05
Moderate (7+ months)	85 (18.9)	70 (18.8)	
Technology Proficiency Level			
Mean ± SD	4.31 ± 1.22	4.36 ± 1.19	t(448) = -0.71, p = .478
Range	1-7	1-7	d = 0.04
Medical Technology Familiarity			
Mean ± SD	3.89 ± 1.34	3.94 ± 1.31	t(448) = -0.67, p = .503
Range	1-7	1-7	d = 0.04
Individual Difference Measures			
Prior Technological Experience Scale			
Mean ± SD	4.18 ± 1.28	4.22 ± 1.25	t(448) = -0.58, p = .562
Range	1-7	1-7	d = 0.03
Learning Goal Orientation			
Mastery Orientation (Mean ± SD)	5.78 ± 0.91	5.81 ± 0.89	t(448) = -0.62, p = .535
Performance Orientation (Mean ± SD)	4.12 ± 1.43	4.09 ± 1.41	t(448) = 0.39, p = .697

Note. *Comprehensive attrition analysis indicates no significant demographic differences between retained and lost participants across all measured variables, supporting missing-at-random assumptions for subsequent analytical procedures. Effect sizes: d = Cohen's d for continuous variables; ϕ = phi coefficient for 2×2 contingency tables; Cramer's V for larger contingency tables. Technology proficiency and familiarity assessed on 7-point Likert scales (1 = very low proficiency/familiarity, 7 = very high proficiency/familiarity). Learning goal orientations assessed using validated psychometric instruments with established reliability and validity evidence. Missing data for any variable $\leq 1.8\%$ across all assessment domains. Statistical power for attrition analyses ranged from .82 to .95 for detecting medium effect sizes ($d \geq 0.30$) at $\alpha = .05$, ensuring adequate sensitivity for identifying systematic attrition patterns.*

Measures

All instruments employed 7-point Likert response scales to optimise measurement precision and distributional variability. To mitigate threats from common method variance (CMV), we implemented a multi-faceted methodological strategy encompassing both procedural and statistical safeguards. Procedurally, temporal separation across the three measurement waves reduced consistency motifs and retrospective recall biases, while psychological separation techniques—including varied response anchors, counterbalanced item sequencing, and the insertion of methodologically unrelated buffer items—diminished respondent hypothesis-guessing and socially desirable responding. Conceptually, we maintained strict theoretical and operational differentiation between AI-assisted diagnosis participation (operationalised as behavioural engagement frequency and depth) and AI literacy (operationalised as cognitive competencies in algorithmic comprehension and evaluation), ensuring minimal construct conflation. Statistically, Harman's single-factor test indicated that no single factor accounted for the majority of covariance (single-factor variance explained = 31.2%), and a confirmatory single-factor model demonstrated substantially degraded fit (CFI = .67, RMSEA = .14) relative to the hypothesised multi-factor measurement model, collectively suggesting that CMV does not pose a serious threat to the validity of the observed associations.

AI-Assisted Diagnosis Participation (Behavior)

The AI-Assisted Diagnosis Participation Scale (AIDPS-15) represents a comprehensive measure integrating behavioral frequency, cognitive engagement depth, and systematic utilization dimensions based on established technology acceptance and engagement frameworks. The 15-item instrument employs 7-point Likert scaling with behavioral anchoring, incorporating items assessing participation frequency ("How often do you actively engage with AI diagnostic systems during clinical activities?"), analytical depth ("To what extent do you critically evaluate AI diagnostic recommendations before clinical application?"), and systematic integration ("How

frequently do you combine AI insights with clinical reasoning in diagnostic decision-making?"). Items were rated from 1 ("never/strongly disagree") to 7 ("very often/strongly agree"), with higher scores indicating more active and reflective participation in AI-assisted diagnosis.

The scale showed excellent internal consistency across waves (Cronbach's $\alpha = .89-.93$) and good temporal stability (test-retest $r = .81$ between adjacent waves). Confirmatory factor analysis supported a three-factor structure corresponding to the frequency, depth, and integration dimensions. Convergent validity was indicated by moderate correlations with faculty ratings of student AI engagement ($r = .64$) and peer assessments of technological participation ($r = .58$). Discriminant validity was supported by weaker correlations with general technology enthusiasm measures ($r = .23$), suggesting that the scale captures participation in clinical AI use rather than general ICT liking.

AI Literacy (Competence)

AI literacy was conceptualised as a competence distinct from usage behaviour. It was measured using an adapted version of the Medical Artificial Intelligence Readiness Scale [12], aligned with core AI literacy components described by Long and Magerko [11]. Items assessed students' perceived ability to understand, evaluate, and work with AI outputs in clinical contexts. Example items include "I can identify potential biases in the data used by the AI model", "I understand the limitations of the AI system's probabilistic reasoning", and "I can explain why an AI recommendation may not be appropriate for a specific patient". Responses were given on a 7-point scale from 1 ("strongly disagree") to 7 ("strongly agree").

Internal consistency of the adapted AI literacy scale was high across waves ($\alpha \geq .90$). To address concerns about construct overlap with AI-assisted diagnosis participation, we conducted a discriminant validity analysis. The square root of the Average Variance Extracted (AVE) for AI literacy (0.84) exceeded its correlation with AI-assisted diagnosis participation ($r = .64$), providing statistical evidence that the two constructs are related but distinct (behaviour vs competence).

Medical Critical Thinking

Medical critical thinking was assessed using the Medical Critical Thinking Scale-Revised (MCTS-R-18), a domain-adapted version of established critical thinking assessments grounded in Facione's critical thinking taxonomy[13]. The scale comprises 18 items capturing analytical reasoning (for example, "I systematically evaluate the evidence before reaching diagnostic conclusions"), inference (for example, "I can identify logical relationships between clinical findings"), and metacognitive regulation (for example, "I monitor my reasoning processes during diagnostic decision-making"). Items were rated from 1 ("strongly disagree") to 7 ("strongly agree"), with higher scores indicating stronger critical thinking dispositions in medical contexts.

The MCTS-R-18 demonstrated good internal consistency across waves ($\alpha = .87-.91$). Previous research has supported its convergent and discriminant validity in medical education contexts, and in the present sample, item distributions and factor loadings were consistent with prior validation studies.

Individual Difference Measures

Two baseline individual difference variables were assessed as potential moderators of the longitudinal associations: prior technological experience and learning goal orientation. Prior technological experience was measured using a 6-item scale assessing baseline competence with medical information systems, previous exposure to AI tools in clinical or educational settings, and general ICT proficiency. Items were rated on a 7-point scale, and higher scores indicated more extensive and confident use of technology.

Learning goal orientation was measured using an 8-item instrument adapted from Dweck and Leggett's achievement goal framework. Items distinguished mastery orientation (for example, "I prefer challenging learning opportunities that help me develop diagnostic skills") from performance orientation (for example, "I prefer clinical situations where I can demonstrate my competence to supervisors"). The instrument employed forced-choice wording to reduce social desirability bias. In this study, mastery and performance orientation scores were treated as continuous indices. Both individual difference measures showed acceptable reliability ($\alpha \geq .83$) and temporal stability, with test-retest correlations ranging from .76 to .82 across measurement occasions. Detailed item content and additional psychometric information for all scales are provided in the Supplementary Table S1. These individual difference measures were administered at all three waves to examine temporal stability; however, only T_0 scores were used as time-invariant moderators in the longitudinal structural models.

Procedure

Ethical approval for this study was obtained from the Ethics Committee of Guangxi Orthopedic Hospital (Approval No. 20221212). All procedures were conducted in accordance with the Declaration of Helsinki and local institutional guidelines. Students were informed about the study aims and procedures during scheduled teaching sessions and were invited to participate voluntarily. Written informed consent was obtained from all participants before data collection.

Questionnaires were administered online via a secure institutional survey platform at each of the three time points. To minimise social desirability bias and perceived coercion, students were assured that their responses would remain confidential, would not be shared with clinical supervisors in an identifiable form, and would have no impact on course grades or clinical performance evaluations. Participants could withdraw from the study at any time without penalty.

Data Analysis

Data analyses were conducted using Mplus version 8.4. Reporting of this observational cohort study follows the STROBE guidelines for longitudinal observational research. Descriptive statistics and bivariate correlations were calculated for all variables at each wave. Confirmatory factor analyses were conducted to evaluate the measurement structure of the multi-item scales, and longitudinal measurement invariance (configural, metric, and scalar) was tested to ensure that construct scores were comparable over time.

The primary analysis tested a three-wave cross-lagged panel mediation model in which AI-assisted diagnosis participation at each wave predicted subsequent AI literacy and medical critical thinking, while controlling for stability paths and concurrent associations among constructs [28]. Baseline age, gender, and grade point average (GPA) were included as covariates predicting the endogenous variables to reduce confounding. Longitudinal mediation was examined by estimating indirect paths from baseline AI-assisted diagnosis participation (T_0) to critical thinking at T_2 via AI literacy at T_1 .

To handle missing data arising from attrition and item non-response, we used Full Information Maximum Likelihood (FIML) estimation under a Missing at Random assumption [29-30]. The inclusion of baseline demographic and academic variables as auxiliary variables in the FIML estimation was intended to improve the plausibility of this assumption and reduce bias associated with selective attrition, auxiliary variables included age, gender, institution type, baseline GPA, and baseline levels of AI participation, AI literacy, and medical critical thinking [29-30]. Item-level missingness was below commonly recommended thresholds for FIML in longitudinal structural equation modelling.

Primary mediation analyses employed Monte Carlo confidence intervals with 10,000 resamples to estimate the precision of indirect effects. To examine whether the indirect pathways differed across learner profiles, we conducted multi-group analyses in which the sample was divided by median splits on prior technological experience, mastery orientation, and performance orientation [31]. In these secondary analyses, equality constraints on the key structural paths were compared with unconstrained models using χ^2 difference tests to evaluate moderation. As a robustness check, sensitivity analyses treating moderators as continuous variables were also conducted; these yielded similar patterns of effects and are reported in the Supplementary Material.

Model fit was evaluated using multiple indices: χ^2 statistics, the Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). CFI and TLI values ≥ 0.95 , RMSEA ≤ 0.06 , and SRMR ≤ 0.08 were considered indicative of good fit, while recognising that rigid adherence to fixed cutoffs is not recommended and that model evaluation should consider the pattern of indices, sample size, and model complexity [28,30,31].

Methodologically, we employed both median-split and continuous-variable approaches to examine moderation effects. The primary analyses used median splits to facilitate clinical interpretability

and to identify distinct learner profiles that may benefit from differentiated instructional approaches. However, acknowledging concerns about statistical power loss and arbitrary boundary effects associated with dichotomisation [38], we conducted supplementary latent moderated structural equations analyses treating all moderators as continuous variables. The convergence of findings across these complementary approaches—with continuous-variable interaction effects reaching significance and conditional indirect effects closely approximating median-split estimates—strengthens confidence in the robustness of the moderation patterns observed. The median-split presentations in the main text serve pedagogical and practical purposes, while the continuous-variable results in Supplementary Table S3 provide more statistically rigorous estimates for methodologically oriented readers.

Data availability

The datasets generated and analysed during the current study are not publicly available due to institutional data protection regulations, but de-identified data may be made available from the corresponding author on reasonable request and with approval from the relevant institutional bodies.

Code availability

All custom code used in this study is available from the corresponding author upon reasonable request. The longitudinal structural equation models were estimated in Mplus (Muthén & Muthén) using robust maximum likelihood estimation with full information maximum likelihood (FIML) for handling missing data. Descriptive and supplementary analyses were conducted in R (R Foundation for Statistical Computing). The Mplus syntax files and R scripts specify all variables, model parameters, and options used to generate and analyse the current datasets.

Acknowledgements

This work was supported by the 2021 Guangxi Philosophy and Social Science Planning Research Project (No. 21FKS028), the Health and Emergency Skills Training Center of Guangxi (No. HESTCG202304), and the 2022 Guangxi Higher Education Undergraduate Teaching Reform Project (No. 2022JGZ155). The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of the manuscript.

We are grateful to all participating students and clinical teaching staff for their time and support in completing the surveys. We also thank the administrative teams at the participating institutions for their assistance with coordinating data collection.

Authors' contributions

Conceptualization: Y.X., D.Y., L.LH.

Methodology: Y.X., D.Y., L.LH.

Formal analysis: Y.X., D.Y., L.LH.

Investigation: Y.X., D.Y., L.MY., L.SR., L.LH.

Data curation: Y.X., D.Y., L.LH.

Writing – original draft: Y.X., D.Y., L.LH.

Writing – review & editing: Y.X., D.Y., L.LH., L.MY., L.SR.

Funding acquisition: Y.X., D.Y., L.LH.

Resources: L.LH., L.MY., L.SR.

Supervision: D.Y., L.LH.

Project administration: D.Y., L.LH.

Competing interests

The authors declare no competing financial or non-financial interests. The authors have no patents or patent applications related to this work, and no author is associated with npj Digital Medicine as an Editorial Board Member, Collection Editor, or Guest Editor.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors did not use generative AI or AI-assisted technologies at any stage of the study design, data analysis, or manuscript preparation.

References

- [1] Topol, E. J. High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56(2019).
- [2] Civaner, M. M., Uncu, Y., Bulut, F., Chalil, E. G., & Tatli, A. Artificial intelligence in medical education: A cross-sectional needs assessment. *BMC Medical Education*, 22(1), 772(2022).
- [3] Thesen, T., & Park, S. H. A generative AI teaching assistant for personalized learning in medical education. *npj Digital Medicine*, 8(1), 627(2025).
- [4] Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ digital medicine*, 3(1), 17(2020).
- [5] McCoy, L. G., Nagaraj, S., Morgado, F., Harish, V., Das, S., & Celi, L. A. What do medical students actually need to know about artificial intelligence?. *NPJ digital medicine*, 3(1), 86 (2020).
- [6] Gaube, S., Suresh, H., Raue, M., Merritt, A., Berkowitz, S. J., Lerner, E., . & Ghassemi, M. Do as AI say: susceptibility in deployment of clinical decision-aids. *NPJ digital medicine*, 4(1), 31(2021).
- [7] Gerlich, M. AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1), 6 (2025).

- [8] Cabral, B. P., Braga, L. A. M., Conte Filho, C. G., & Mota, F. Future use of AI in diagnostic medicine: 2-wave cross-sectional survey study. *Journal of Medical Internet Research*, 27, e53892(2025).
- [9] Gonsalves, C. Generative AI's impact on critical thinking: Revisiting Bloom's taxonomy. *Journal of Marketing Education*, 46(2), 150–165(2024).
- [10] Chan, K. S., & Zary, N. Applications and challenges of implementing artificial intelligence in medical education: Integrative review. *JMIR Medical Education*, 5(1), e13930(2019).
- [11] Long, D., & Magerko, B. What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). ACM(2020).
- [12] Karaca, O., Caliskan, S. A., & Demir, K. Medical artificial intelligence readiness scale for medical students (MAIRS-MS) – development, validity and reliability study. *BMC Medical Education*, 21, 112(2021).
- [13] Facione, P. A. *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. American Philosophical Association(1990).
- [14] Fogel, A. L., & Kvedar, J. C. Artificial intelligence powers digital medicine. *NPJ digital medicine*, 1(1), 5(2018).
- [15] Masters, K. Ethical use of artificial intelligence in health professions education: AMEE Guide No. 158. *Medical Teacher*, 45(6), 574–584(2023).
- [16] Sapci, A. H., & Sapci, H. A. Artificial intelligence education and tools for medical and health informatics students: Systematic review. *JMIR Medical Education*, 6(1), e19285(2020).
- [17] Dweck, C. S. Motivational processes affecting learning. *American Psychologist*, 41(10), 1040–1048(1986).
- [18] Beik, A., & Cho, Y. Effects of goal orientation on online learning: A meta-analysis of differences in Korea and US. *Current Psychology*, 43(2), 1496-1506(2024).
- [19] Pallant, J. L., Blijlevens, J., Campbell, A., & Jopp, R. Mastering knowledge: the impact of generative AI on student learning outcomes. *Studies in Higher Education*, 1-22(2025).
- [20] Wang, S., Shi, G., Lu, M., Lin, R., & Yang, J. Determinants of active online learning in the smart learning environment: An empirical study with PLS-SEM. *Sustainability*, 13(17), 9923(2021).

- [21] Yang, X., Deng, S., Hu, W., & Deng, Y. Associations between social media use and academic adjustment among university students: An integrated variable-centered and person-centered analytical framework. *Problems of Education in the 21st Century*, 83(3),383–402(2025).
- [22] Xin, Y., Shusheng, D., Weina, H., & Yan, D. How does digital connection shape cultural adaptation? The impact of social media use on cross-cultural adaptation of international students in China. *Journal of International Students*, 15(9), 1-26(2025).
- [23] Cheung, M. W. L., & Lau, R. S. Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models. *Organizational Research Methods*, 11(2), 296-325(2008).
- [24] Preacher, K. J., & Hayes, A. F. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879-891 (2008).
- [25] Sweller, J., van Merriënboer, J. J. G., & Paas, F. Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292(2019).
- [26] Wang, W. Medical education in China: progress in the past 70 years and a vision for the future. *BMC Medical Education*, 21(1), 453(2021).
- [27] Zhang, S., Ganapathy Prasad, P., & Schroeder, N. L. Learning About AI: A Systematic Review of Reviews on AI Literacy. *Journal of Educational Computing Research*, 07356331251342081(2025).
- [28] Wu, W., Carroll, I. A., & Chen, P. Y. A single-level random-effects cross-lagged panel model for longitudinal mediation analysis. *Behavior Research Methods*, 50(5), 2111-2124(2018).
- [29] Enders, C. K. *Applied missing data analysis*. Guilford Publications(2022).
- [30] Velasquez, G., & Zhang, Q. Cross-lagged Panel Mediation Models with Latent Constructs: Specification and Estimation. *Multivariate Behavioral Research*, 55(1), 142-143(2020).
- [31] Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., .& Diener, E. From data to causes I: Building a general cross-lagged panel model (GCLM). *Organizational Research Methods*, 23(4), 651-687(2020).
- [32] Kolachalama, V.B. and Garg, P.S., Machine learning and medical education. *NPJ digital medicine*, 1(1), 54(2018).

- [33] Fiedler, K., Schott, M. and Meiser, T.,. What mediation analysis can (not) do. *Journal of Experimental Social Psychology*, 47(6), 1231-1236(2011).
- [34] Fiedler, K., Harris, C. and Schott, M.,. Unwarranted inferences from statistical mediation tests—An analysis of articles published in 2015. *Journal of Experimental Social Psychology*, 75, 95-102(2018).
- [35] Klein, A. and Moosbrugger, H., Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, 65(4), 457-474(2000).
- [36] Maslowsky, J., Jager, J. and Hemken, D., Estimating and interpreting latent variable interactions: A tutorial for applying the latent moderated structural equations method. *International journal of behavioral development*, 39(1), 87-96(2015).
- [37] Preacher, K.J., Curran, P.J. and Bauer, D.J.,. Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of educational and behavioral statistics*, 31(4), 437-448(2006).
- [38] MacCallum, R.C., Zhang, S., Preacher, K.J. and Rucker, D.D., 2002. On the practice of dichotomization of quantitative variables. *Psychological methods*, 7(1), p.19.