Full paper

Sains
Humanika

Beyond Cognitive Offloading: AI Reliance and Perceived Higher-Order Thinking Skills among UNIMAS Undergraduates

Cheryl Jia-Ying Chung, Kee-Man Chuah*

Faculty of Education, Language and Communication, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

*Corresponding author: kmchuah@unimas.my

Article history: Received:27 March 2026 Received in revised form:04 June 2026 Accepted: 05 June 2026 Published online: 17 September 2026

Abstract

The increasing use of artificial intelligence (AI) tools in higher education has prompted widespread concern that student reliance on these technologies may erode higher-order thinking skills (HOTS). However, much of the existing evidence draws on Western contexts and relies on deficit-oriented assumptions about AI use. This mixed-methods study investigates the relationship between reliance on AI and perceived higher-order thinking skills among undergraduate students at Universiti Malaysia Sarawak (UNIMAS), with a comparative focus on STEM and non-STEM disciplines. Quantitative data were collected from 100 undergraduates using a structured questionnaire, while qualitative data were collected from six purposively selected participants through semi-structured interviews. Contrary to the dominant cognitive offloading narrative, a strong positive correlation was found between AI reliance and self-reported engagement in higher-order thinking ($\rho = .577$, $p < .001$), with no significant disciplinary differences. Qualitative findings revealed that students approached AI with verification strategies, critical awareness, and reflective judgement rather than passive dependence. These results suggest that the relationship between AI use and cognitive engagement is more nuanced than prevailing literature implies, and that contextual factors such as evaluative habits and pedagogical culture may be more influential than disciplinary background in shaping how students think with AI

Keywords: artificial intelligence reliance, cognitive offloading, higher-order thinking skills, STEM and non-STEM, Malaysian higher education

Abstrak

Peningkatan penggunaan alat kecerdasan buatan (AI) dalam pendidikan tinggi telah mencetuskan kebimbangan meluas bahawa kebergantungan pelajar terhadap teknologi ini boleh menghakis kemahiran berfikir aras tinggi (KBAT). Walau bagaimanapun, kebanyakan bukti sedia ada adalah berdasarkan konteks Barat dan bergantung pada andaian berorientasikan defisit tentang penggunaan AI. Kajian kaedah campuran ini menyiasat hubungan antara kebergantungan terhadap AI dan kemahiran berfikir aras tinggi yang ditanggap dalam kalangan mahasiswa prasiswazah di Universiti Malaysia Sarawak (UNIMAS), dengan tumpuan perbandingan antara disiplin STEM dan bukan STEM. Data kuantitatif dikumpulkan daripada 100 orang mahasiswa prasiswazah melalui soal selidik berstruktur, manakala data kualitatif diperoleh daripada enam orang peserta yang dipilih secara bertujuan melalui temu bual separa berstruktur. Bertentangan dengan naratif pemunggahan kognitif yang dominan, korelasi positif yang kuat ditemui antara kebergantungan terhadap AI dan penglibatan berfikir aras tinggi yang dilaporkan sendiri ($\rho = .577$, $p < .001$), tanpa perbezaan yang signifikan. Dapatan kualitatif mendedahkan bahawa pelajar menggunakan AI dengan strategi pengesahan, kesedaran kritikal, dan pertimbangan reflektif, bukannya kebergantungan pasif. Keputusan ini mencadangkan bahawa hubungan antara penggunaan AI dan penglibatan kognitif adalah lebih bernuansa daripada apa yang tersirat dalam kajian lepas, dan faktor kontekstual seperti amalan penilaian serta budaya pedagogi mungkin lebih berpengaruh daripada latar belakang disiplin dalam membentuk cara pelajar berfikir dengan AI.

Kata kunci: kebergantungan terhadap kecerdasan buatan, pemunggahan kognitif, kemahiran berfikir aras tinggi, STEM dan bukan STEM, pendidikan tinggi Malaysia

■1.0 INTRODUCTION

A dominant narrative has emerged in the educational technology literature: that student reliance on artificial intelligence (AI) tools poses a threat to higher-order thinking skills (HOTS). This concern is not without basis. Several studies have reported negative associations between frequent AI use and critical thinking (Gerlich, 2025), problem-solving (Çela et al., 2024), and independent reasoning (Li et al., 2025). The theoretical underpinning is cognitive offloading, the idea that delegating cognitive tasks to external tools reduces the mental effort invested in deep processing, and that this reduction has consequences for analytical and evaluative reasoning.

Nevertheless, this narrative warrants scrutiny. Much of the evidence is drawn from Western contexts with distinct pedagogical cultures, and the assumption that AI use necessarily displaces cognitive effort overlooks the possibility that students may use AI in ways that complement or even stimulate critical engagement. Risko and Gilbert (2016) noted that cognitive offloading is not inherently detrimental; its effects depend on how, when, and why individuals offload. A student who uses ChatGPT to generate a first draft and then critically evaluates it is engaging in a fundamentally different cognitive process from one who submits the output unexamined.

Despite the potential for disciplinary differences in how students engage with AI, comparative research remains sparse. Parviz (2024) compared STEM and non-STEM instructors' attitudes toward AI integration but did not examine student perspectives. The question of whether disciplinary training shapes how students think with AI, rather than merely how often they use it, remains largely unanswered. This gap is particularly significant in the Malaysian higher education context, where AI adoption is accelerating but empirical evidence on its cognitive implications for students remains limited (Selvanathan & Narayanan, 2024).

The rapid integration of AI into higher education has been well documented (Adams & Chuah, 2022; Chen et al., 2020; Limna et al., 2022). ChatGPT has seen widespread adoption across university contexts globally, transforming how students approach tasks ranging from literature searches to essay writing and data analysis. However, the speed of adoption has outpaced empirical understanding of its cognitive consequences. Most existing studies have examined AI use in isolation, treating it as an independent variable that either helps or hinders learning, without attending to the mediating role of students' own evaluative practices. This study takes a different approach by foregrounding students' perceptions and practices, asking not simply whether AI use correlates with cognitive outcomes, but how students describe their own cognitive engagement when using AI tools.

In Malaysia, the Ministry of Higher Education has increasingly emphasised the development of 21st-century skills, including critical thinking and digital literacy, as central to the national higher education agenda (Khan et al., 2022; Ministry of Higher Education Malaysia, 2025). This policy environment creates a distinctive context for studying AI's cognitive effects, as students may be exposed to explicit pedagogical messaging emphasising the importance of critically evaluating information. The extent to which this institutional emphasis translates into actual evaluative practices (Chuah & Sumintono, 2024) when students interact with AI tools is an empirical question that the present study seeks to address.

This study addresses these gaps by examining the relationship between reliance on AI and perceived higher-order thinking skills among undergraduates, with attention to whether disciplinary background (STEM versus non-STEM) influences this relationship. The objectives are as follows:

i. To determine the relationship between students' perceived AI reliance and their higher-order thinking skills.
ii. To compare AI reliance and its perceived effects on higher-order thinking skills between STEM and non-STEM students.
iii. To explore how STEM and non-STEM students perceive the role of AI in influencing their higher-order thinking skills.

2.0 LITERATURE REVIEW

2.1 AI Reliance and Cognitive Offloading

Reliance on artificial intelligence refers to the tendency to accept and act on AI-generated recommendations, particularly when they diverge from one's own judgment (Guo et al., 2024). This construct is closely related to cognitive offloading, defined as the use of external resources to reduce a task's cognitive demands (Risko & Gilbert, 2016). In the context of higher education, cognitive offloading through AI may manifest as students relying on AI-generated summaries rather than reading primary sources or accepting AI-suggested answers without verification.

The empirical evidence on the cognitive consequences of AI reliance is, however, more mixed than the prevailing narrative suggests. Gerlich (2025) reported a significant negative association between AI usage frequency and critical thinking in a UK sample of 666 participants, identifying cognitive offloading as a key mediating factor. Çela et al. (2024) found no difference in critical thinking between students with and without prior AI exposure but did identify a negative correlation between heavy reliance on AI and problem-solving ability. These findings have been widely cited as evidence that AI undermines higher-order thinking.

The concept of cognitive offloading has gained renewed attention in the AI era, but it predates generative AI by decades. Risko and Gilbert (2016) defined it as the use of physical actions or external resources to alter the information processing requirements of a task. Their review concluded that offloading is a strategic, context-dependent behaviour, not an inherently harmful one. Individuals offload when the cost of internal processing exceeds the cost of external support, and this trade-off can be rational. The question, then, is not whether students offload cognitive work to AI, which they almost certainly do, but whether the specific forms of offloading that AI enables are detrimental to the development of higher-order thinking. This distinction has been largely absent from the educational technology discourse, which has tended to treat all forms of AI-assisted cognition as equivalent.

In addition, Obenza et al. (2024) found that university students perceived ChatGPT as useful for academic tasks but expressed concerns about accuracy and over-dependence. Shakil and Siddiq (2024) highlighted institutional concerns about assessment integrity arising from AI use among ESL learners at the graduate level. Zhou et al. (2024), in a mixed-methods study of UK university students, identified common academic uses of AI alongside concerns related to diminished independent thinking. These studies collectively paint a picture of students who are aware of AI's limitations even as they rely on it, suggesting a more reflective engagement than the term "offloading" might imply.

However, several studies complicate this conclusion. Kofahi and Husain (2025) found that AI-supported constructivist learning environments strengthened higher-order cognitive functions, suggesting that pedagogical design mediates the relationship between AI use and cognitive outcomes. Huesca et al. (2024) demonstrated that integrating ChatGPT into a flipped learning approach improved conceptual understanding among 356 engineering undergraduates. Kosar et al. (2024) observed that students who engaged critically with ChatGPT for code optimisation maintained their analytical reasoning without any decline in overall performance. These contrasting findings suggest that the impact of AI on thinking is not determined by the technology itself, but by the conditions of its use.

2.2 Higher-Order Thinking Skills in an AI-Mediated Context

Higher-order thinking skills encompass the cognitive processes of analysis, evaluation, and synthesis that extend beyond factual recall (Kofahi & Husain, 2025). The concern that AI may compromise these skills rests on the assumption that students who receive ready-made answers are deprived of the cognitive friction necessary for deep learning. There is intuitive appeal to this argument, but it is worth noting that similar concerns have accompanied every major educational technology, from calculators to search engines, and the evidence for lasting cognitive harm has been less clear-cut than initial fears suggested.

A more productive framing may be to ask not whether AI undermines HOTS, but under what conditions students maintain or develop HOTS while using AI. Essien et al. (2024) found that AI tools supported lower-level cognitive tasks among non-STEM postgraduate students but produced limited gains at higher-order levels, suggesting that the cognitive ceiling of AI assistance may differ by task type rather than by discipline. Valeri et al. (2025) reported similar patterns among STEM students, with ChatGPT proving useful for conceptual understanding but unreliable for tasks requiring mathematical precision. These findings point to a more granular picture in which AI's cognitive impact varies by the nature of the task, the student's existing competence, and the degree of critical engagement applied to AI outputs.

The measurement of higher-order thinking skills itself presents methodological challenges that are often underacknowledged in the AI-in-education literature. Self-report measures of HOTS capture students' perceptions of their own critical thinking, which may diverge substantially from actual cognitive performance. As noted by Pennycook et al. (2017), the Dunning-Kruger effect, whereby individuals with limited competence overestimate their abilities, and social desirability bias, whereby respondents report what they believe is expected of them, can both inflate self-reported HOTS scores. Performance-based assessments, such as those used by Gerlich (2025), may yield different patterns. This methodological tension is not merely a limitation to be noted in passing; it is a substantive analytical issue that shapes the interpretation of any study, including the present one, that relies on self-reported cognitive engagement.

Yatani et al. (2024) proposed the concept of AI as "extraherics," arguing that AI tools can foster higher-order thinking when they are positioned as cognitive partners rather than cognitive replacements. Their framework suggests that the key variable is not the frequency or intensity of AI use but the quality of the human-AI interaction, specifically whether the interaction involves reflection, evaluation, and synthesis by the human user. This perspective aligns with constructivist learning theory, which holds that knowledge is actively constructed through engagement with material rather than passively received.

2.3 Disciplinary Differences and the Malaysian Context

The STEM/non-STEM distinction in AI research often proceeds from the assumption that disciplinary training produces fundamentally different cognitive orientations toward technology. STEM students, the reasoning goes, are more accustomed to precision and verification, while non-STEM students may approach AI as a generative or interpretive tool (Parviz, 2024). However, the empirical basis for this distinction is thin. Most comparative studies have focused on instructors rather than students, and the few student-focused studies have produced inconclusive results.

In the Malaysian context, several factors may shape how students engage with AI in ways that differ from those in Western settings. English is typically a second language for most students, meaning AI tools may serve a compensatory linguistic function alongside their cognitive role (Lijie et al., 2024). Institutional cultures at Malaysian universities may also foster evaluative norms around technology use. Wilson and Narasuman (2020) noted that Malaysian higher education has placed increasing emphasis on critical thinking as an explicit learning outcome, which may prompt students to scrutinise AI outputs more closely than their counterparts in settings where critical thinking is less formally embedded in curricula. These contextual factors suggest that findings from Western studies may not transfer straightforwardly to the Malaysian setting, strengthening the case for locally situated research.

Hasin et al. (2025) found a dual pattern in the Malaysian context, whereby AI-assisted learning tools enhanced efficiency and comprehension but simultaneously raised concerns about over-reliance and diminished independent thinking. Amdan et al. (2025) reported that STEM students in Malaysian TVET institutions used AI tools for accuracy-oriented tasks yet expressed concerns about the erosion of critical thinking. These Malaysian-specific findings suggest a nuanced landscape in which students are neither uncritical adopters nor wholesale rejectors of AI, but rather strategic users who negotiate the tension between convenience and cognitive independence.

Selvanathan and Narayanan (2024) examined whether ChatGPT represents an opportunity or a threat to the Malaysian education system, concluding that the answer depends heavily on how AI is integrated into pedagogical practice. Their study highlighted the need for empirical research that moves beyond binary framings of AI as either beneficial or harmful and instead examines the conditions under which AI use supports or undermines specific learning outcomes. The present study responds to this call by investigating the relationship between reliance on AI and perceived HOTS in a specific Malaysian institutional context.

■3.0 METHODOLOGY

3.1 Research Design

This study employed a mixed-methods descriptive research design. Quantitative data addressed the relationship between AI reliance and higher-order thinking skills (RO1) and disciplinary differences (RO2), while qualitative data captured students' perceptions of AI use in greater depth (RO3). Integrating both strands enabled triangulation, as quantitative findings could be cross-checked, contextualised, and enriched by qualitative accounts (Morse, 2016).

3.2 Context and Participants

The study was conducted at Universiti Malaysia Sarawak (UNIMAS) with undergraduate students from a range of academic programmes. A stratified sampling approach ensured balanced representation across STEM and non-STEM disciplines. A total of 100 undergraduates participated in the survey, comprising equal numbers from each disciplinary group. Of the 100 participants, 66 (66%) were female and 34 (34%) were male. The majority of respondents were Year 2 students (78%), followed by Year 1 students (15%), Year 4 (5%), and Year 3 (2%). This concentration in the first two years of study is a limitation, as it may not capture the perspectives of more advanced students whose relationship with AI could differ due to greater disciplinary socialisation. Table 1 shows the distribution of participants by their respective faculties and discipline category.

Table 1 Distribution of Participants by Faculties and Discipline Category

Discipline	Faculty	n
STEM	Faculty of Computer Science and Information Technology (FCSIT)	22
	Faculty of Engineering (FENG)	19
	Faculty of Resource Science and Technology (FSTS)	5
	Faculty of Medicine and Health Sciences (FMHS)	4
	Total STEM	50
Non-STEM	Faculty of Economics and Business (FEB)	17
	Faculty of Cognitive Sciences and Human Development (FCSHD)	14
	Faculty of Education, Language, and Communication (FELC)	13
	Faculty of Applied and Creative Arts (FACA)	3
	Faculty of Social Sciences and Humanities (FSSH)	3
	Faculty of Built Environment (FBE)	0
	Total Non-STEM	50

The sample size of 100 participants, while modest, is consistent with comparable studies in the Malaysian higher education context (Hasin et al., 2025; Selvanathan & Narayanan, 2024) and exceeds the minimum threshold for non-parametric correlation and group comparison analyses. However, the equal split between STEM ($n = 50$) and non-STEM ($n = 50$) was achieved through stratified sampling rather than proportional representation, meaning the sample does not reflect the actual enrolment ratio across faculties at UNIMAS. The uneven distribution within each disciplinary category should also be noted, as the STEM group is dominated by computing and engineering students, which may not represent the full range of STEM disciplines. The near-exclusive reliance on ChatGPT (86%) is a further constraint, as findings may not generalise to students using other AI tools with different affordances and interaction paradigms.

ChatGPT was the overwhelmingly dominant AI tool (86%), with Microsoft Copilot and Google Gemini each at 6%, and all other tools at 1% or below. Most respondents (85%) did not subscribe to paid AI tools. For the qualitative component, six participants were purposively selected, three from STEM and three from non-STEM faculties.

3.3 Instruments

The instrument was a structured questionnaire comprising twenty-two items across four sections, administered in English. A four-point Likert scale (1 = Strongly Disagree to 4 = Strongly Agree) was used to measure AI reliance, cognitive offloading tendencies, and self-reported higher-order thinking skills. Sections B–D contained sixteen Likert-scale items adapted from Gerlich (2025), with minor modifications to contextualise items for the UNIMAS population. Section B (five items) measured AI reliance for information retrieval and academic decision-making. Section C (five items) measured cognitive offloading tendencies. Section D (six items) assessed self-reported critical thinking. A pilot test was conducted with 14 undergraduate students not included in the final sample to ensure item clarity and relevance (Bowden et al., 2002). The use of a four-point scale, while avoiding a neutral midpoint, may force responses from participants who hold genuinely ambivalent positions. Furthermore, the reliance on self-reported HOTS rather than performance-based assessment is a significant methodological consideration addressed in the Discussion.

For the qualitative component, semi-structured interviews were conducted with six purposively selected participants to ensure representation from both STEM and non-STEM disciplines. The interview protocol comprised six open-ended questions covering two domains: AI reliance (three questions exploring the ways, extent, and perceived efficiency of AI tool use in academic work) and impact on higher-order thinking skills (three questions examining perceived influence on critical thinking, verification practices, and instances of uncritical acceptance of AI outputs). Interviews were conducted in mid-August 2025, either face-to-face or online depending on participant availability, and lasted between 20 and 35 minutes each. All interviews were audio-recorded with participant consent and subsequently transcribed verbatim for analysis.

3.4 Procedure and Data Analysis

The questionnaire was distributed between mid-July and 6 August 2025 via Google Forms through faculty-specific WhatsApp and Telegram groups, with supplementary in-person recruitment. Semi-structured interviews were conducted in mid-August 2025. Quantitative data were analysed using Jamovi (Version 2.3.28). The questionnaire demonstrated good reliability ($\alpha = .798$). The Shapiro–Wilk test indicated non-normal distributions for both main variables ($p < .05$), and non-parametric tests were consequently employed, namely Spearman's rank-order correlation for RO1 and the Mann–Whitney U test for RO2. Correlation coefficients were interpreted as weak (<

.30), moderate (.30–.49), or strong (≥ .50). Two hypotheses were formulated to be tested; $H_0 1$: There is no significant relationship between AI reliance and higher-order thinking skills among UNIMAS students, and $H_0 2$: There are no significant differences in AI reliance and higher-order thinking skills (HOTS) between STEM and non-STEM students. An inductive thematic analysis following Braun and Clarke's (2006) six-phase framework was conducted for RO3.

3.5 Ethics

Participation was voluntary, and informed consent was obtained. Survey responses were anonymous, and interview data were treated confidentially. Validated items, triangulation, and reflexive awareness were employed during qualitative analysis to minimise bias.

■4.0 RESULTS

4.1 Descriptive Statistics

Table 2 presents the descriptive statistics for AI reliance, measured on a four-point Likert scale.

Table 2 Mean and Standard Deviation of AI Reliance (N = 100)

Items	M	SD
I use AI tools such as ChatGPT or Copilot specifically for academic purposes.	3.51	0.63
I believe AI tools help me find academic information more quickly and easily than before.	3.45	0.63
I double-check AI-generated information by comparing it with academic references.	3.39	0.76
Before using AI suggestions, I take time to think about whether they are useful or appropriate.	3.34	0.74
I use AI tools to complete academic tasks more efficiently and meet deadlines.	3.24	0.71
I use AI tools to make decisions related to academic planning and study strategies.	3.13	0.85
I depend on digital tools for managing my academic tasks and accessing learning resources.	3.11	0.65
When I face an academic question, I first look online for answers.	3.09	0.81
I use digital tools to store and organise my academic content (e.g., notes, deadlines).	3.07	0.93
When I need information, I usually ask AI tools or search online rather than trying to remember it.	2.99	0.83

Respondents reported high reliance on AI for academic purposes (M = 3.51, SD = 0.63) and perceived these tools as helpful for locating information efficiently (M = 3.45, SD = 0.63). Notably, items reflecting critical engagement scored almost as highly as those reflecting reliance. Students reported verifying AI outputs against academic sources (M = 3.39, SD = 0.76) and reflecting on the appropriateness of AI suggestions before applying them (M = 3.34, SD = 0.74). This pattern is striking because the AI reliance construct and the verification items are positively correlated within the same respondent pool, suggesting that higher reliance does not preclude evaluative behaviour.

Cognitive offloading items revealed moderate-to-high tendencies toward external information management. Students often rely on digital tools to organise academic materials such as notes and deadlines (M = 3.07, SD = 0.93), and some prefer to seek information online or through AI rather than recall it from memory (M = 2.99, SD = 0.83). These findings reflect a broader shift towards external cognitive support, in which technology assists with information handling and retrieval. The finding that the lowest-scoring item (M = 2.99) still falls close to the "Agree" anchor suggests that cognitive offloading is widespread among this student population, regardless of disciplinary background. Table 3 presents the descriptive statistics for higher-order thinking skills.

Table 3 Mean and Standard Deviation of Higher-Order Thinking Skills (N = 100)

Items	M	SD
Before using any academic information, I make sure it comes from a trustworthy source.	3.48	0.64
I think critically about the ideas behind the information AI tools give me.	3.43	0.64
I think twice before accepting what AI tools say as correct or reliable.	3.41	0.64
I compare insights from AI tools with information from scholarly sources.	3.32	0.63
I reflect on how my assumptions or biases influence my academic decisions.	3.23	0.74
I can identify misleading or biased information in AI-generated content.	3.22	0.69

Students reported moderately high engagement in higher-order cognitive processes. The relatively high and uniform scores (range: 3.22–3.48), combined with low standard deviations, should be interpreted cautiously. This pattern may reflect genuine critical engagement, but it is also consistent with social desirability bias, a well-documented tendency in self-report instruments measuring socially valued competencies.

The finding that source verification (M = 3.48) scored highest among the HOTS items is noteworthy in the Malaysian context, where academic integrity and proper attribution are increasingly emphasised in institutional policies. It is possible that the high scores on verification-related items reflect institutional norms rather than spontaneous critical engagement. Conversely, the relatively lower scores on reflective items, such as questioning personal assumptions (M = 3.23) and identifying bias in AI content (M = 3.22), suggest that the more cognitively demanding aspects of critical thinking, those requiring introspection and evaluative judgement rather than procedural verification, may be less consistently practised. This gradient within the HOTS scores offers a more nuanced picture than the aggregate statistics alone would suggest.

4.2 RO1: Relationship Between AI Reliance and HOTS

The Shapiro–Wilk test confirmed non-normality for both AI Reliance ($W = 0.952$, $p = .001$) and HOTS ($W = 0.892$, $p < .001$). A Spearman's rank-order correlation was therefore conducted. A strong positive correlation was found between AI reliance and higher-order thinking skills ($\rho = .577$, $p < .001$). The null hypothesis (H_01) was rejected. This finding indicates that students who reported greater reliance on AI also reported higher levels of critical and evaluative engagement, a result that runs counter to the cognitive offloading hypothesis in its strongest form.

4.3 RO2: Differences Between STEM and Non-STEM Students

As shown in Table 4, no statistically significant differences were found for AI reliance ($U = 1200$, $p = .730$) or HOTS ($U = 1171$, $p = .584$). The null hypothesis (H_02) was not rejected. Disciplinary background, at least as operationalised through the STEM/non-STEM binary, did not meaningfully differentiate students' reliance on AI or their perceived engagement in higher-order thinking.

Table 4 Mann–Whitney U Tests Comparing AI Reliance and HOTS Between STEM and Non-STEM Students

Variable	Statistic	p
AI Reliance	1200	.730
HOTS	1171	.584

Note. H_a: μ(Non-STEM) $\neq$ μ(STEM)

4.4 RO3: Qualitative Findings on User Perceived Role of AI on HOTS

Thematic analysis revealed five sub-themes regarding AI reliance. Participants across both groups described AI as a valuable academic aid for efficiency. STEM students tended to describe AI as a tool for validation, with S3 noting that "AI can act as a springboard for deeper thought by providing a quick overview." Non-STEM students framed AI as a creative collaborator, with NS4 stating, "I use ChatGPT as a brainstorming partner." However, these differences were matters of emphasis rather than fundamental divergence, which aligns with the non-significant quantitative findings.

The sub-theme AI as a language and communication aid emerged strongly. S5 explained, "Sometimes we already have the idea, but we don't know how to elaborate. So, we just give AI a short draft, and it helps us turn it into complete, proper sentences." This finding is significant because it suggests that, in the Malaysian context, reliance on AI may partly reflect a compensatory linguistic strategy rather than cognitive dependence. Students who use AI to articulate pre-existing ideas in more polished English are engaged in a fundamentally different process from those who rely on AI to generate the ideas themselves.

Regarding impact on higher-order thinking skills, students articulated precisely the kind of metacognitive reasoning that the cognitive offloading literature worries AI will suppress. NS2 described posing the same question to two different tools to evaluate divergent responses, representing a form of evaluative reasoning arguably stimulated by AI rather than diminished by it. S1 described checking, comparing, and deciding before use, a deliberate evaluative sequence that mirrors the analytical processes central to Bloom's higher-order cognitive taxonomy.

At the same time, students were not uncritically positive. S3 warned that "if overused, they can hinder critical thinking by providing ready-made answers," while NS4 described the temptation to accept AI outputs "without deeper evaluation." This dual awareness, appreciating AI's utility while remaining alert to its cognitive costs, suggests a more sophisticated relationship with AI than simple dependence. S3's account of initially trusting AI outputs before encountering errors and subsequently developing verification habits illustrates how experience with AI can itself generate critical thinking, a dynamic that purely quantitative studies may fail to capture.

The theme of human–AI collaboration in learning revealed that some participants used AI collaboratively within group settings. S5 remarked, "Normally, I use AI when our group has no ideas. If people don't know what to do, it's best to ask AI. Then, when AI gives an answer, we can discuss from there." This suggests that, rather than isolating students, AI occasionally served as a conversation starter that stimulated peer discussion and collective reasoning. NS6 explained, "I mainly use ChatGPT for design inspiration; it gives me ideas that I can then adapt," demonstrating how AI's generative capabilities promoted ideational diversity while maintaining the student's role as the final creative decision-maker.

The theme of balancing efficiency with independent thinking encapsulated students' recognition that AI's usefulness must coexist with cognitive effort. NS2 reflected, "It really saves time, but it doesn't always improve learning efficiency," while S3 advised, "It's important to use the saved time for deeper learning and critical analysis, not for cutting corners." These insights underline that while AI enhances productivity, genuine learning requires deliberate reflection and self-regulation. The qualitative findings, taken together, illustrate that both STEM and non-STEM students demonstrated awareness of AI's potential to both aid and impede higher-order thinking, embodying verification, evaluation, and synthesis rather than passive dependence.

■5.0 DISCUSSION AND RECOMMENDATIONS

The most striking finding of this study is the strong positive correlation between AI reliance and perceived higher-order thinking skills ($\rho = .577$, $p < .001$), a result that directly challenges the prevailing narrative that AI use undermines critical thinking. This positive correlation must be interpreted with care. It does not demonstrate that reliance on AI enhances critical thinking, nor does it rule out the possibility that the relationship is an artefact of self-report methodology. Students who are more cognitively engaged may both use AI more frequently (because they are more active learners generally) and rate themselves more highly on critical thinking measures, without the two being causally linked.

Furthermore, self-reported HOTS may reflect what students believe they do, or what they consider socially desirable to report, rather than what they do when engaging with AI. Gerlich (2025), whose instrument was adapted for this study, used performance-based assessments alongside self-report measures and found the opposite pattern, a negative relationship, suggesting that the measurement approach may substantially influence findings.

Despite these caveats, the finding is not without explanatory value. It aligns with studies that have found positive cognitive outcomes when AI is used in structured and evaluative ways (Kofahi & Husain, 2025; Huesca et al., 2024). Kosar et al. (2024) observed that students who engaged critically with ChatGPT for code optimisation maintained their analytical reasoning, and Hasin et al. (2025) found a dual pattern in which AI enhanced efficiency and comprehension while also raising concerns about dependency. The present study's qualitative data provide a mechanism for understanding the positive correlation: students described verification behaviours, cross-referencing strategies, and reflective judgements that constitute genuine higher-order thinking. AI may not be undermining these students' cognitive skills; rather, it may be providing material to think critically about.

The absence of significant disciplinary differences in either AI reliance or HOTS is the study's second notable finding. This null result is consistent with Kosar et al. (2024), who found no impact of ChatGPT use on course achievement across disciplines. The qualitative findings revealed that disciplinary differences were matters of emphasis (STEM students favouring validation, non-STEM students favouring ideation) rather than fundamental divergence. The implication is that the STEM/non-STEM binary, while convenient for research design, may be too coarse a distinction to capture meaningful variation in how students engage cognitively with AI. Digital literacy, individual evaluative habits, and pedagogical context may be more consequential variables, a point also noted by Amdan et al. (2025) in the Malaysian TVET context.

The qualitative finding that AI serves a compensatory linguistic function warrants particular attention, as it suggests a dimension of AI reliance largely absent from Western-centric studies. In contexts where students write academically in a second or third language, the boundary between cognitive support and linguistic support is blurred. A student who uses AI to convert a rough draft into polished academic English may appear highly AI-reliant on a quantitative measure, while in cognitive terms they are doing substantial independent thinking prior to the AI interaction (Barrot, 2023). Future research in multilingual higher education contexts should develop instruments to disentangle reliance on linguistic AI from reliance on cognitive AI, as the two may have quite different implications for higher-order thinking.

Within the Malaysian context, two factors may help explain why these findings diverge from Western studies reporting cognitive decline. First, the compensatory linguistic function of AI for students writing in English as a second language may mean that reliance scores partly capture language support rather than cognitive dependence. Lijie et al. (2024) found that perceived ease of use negatively affected critical thinking dispositions among Malaysian students, but perceived usefulness strengthened engagement, suggesting that the purpose of AI use matters more than its frequency. Second, the increasing institutional emphasis on critical thinking as an explicit learning outcome in Malaysian higher education (Wilson & Narasuman, 2020) may create evaluative norms that shape how students approach AI outputs.

It is also worth considering whether the positive correlation may partly reflect measurement confounds. The researchers admitted that several items in the AI reliance scale (e.g., "I double-check AI-generated information by comparing it with academic references") describe behaviours that are themselves forms of higher-order thinking. If students who verify AI outputs score highly on both the reliance and HOTS scales, the positive correlation may partly capture an overlap in what the two constructs measure rather than a genuine relationship between reliance and cognitive engagement. Future instrument development should address this potential confound by ensuring clearer construct separation between AI use patterns and indicators of cognitive engagement.

The qualitative data also raise questions about the role of metacognitive awareness in mediating the relationship between AI use and cognitive outcomes. Students who can articulate the risks of AI dependence, as several participants in this study did, may be precisely the ones best positioned to mitigate them. As echoed by Lee et al. (2026), if metacognitive awareness moderates the relationship between AI reliance and HOTS, then educational interventions aimed at developing students' metacognitive skills may be more effective than restricting AI use. This hypothesis could not be tested with the present study's design, but it offers a productive direction for future research.

The finding that the STEM/non-STEM binary does not capture meaningful variation in AI-related cognitive outcomes has implications for how researchers design comparative studies. Rather than relying on broad disciplinary categories, future research might examine differences among specific academic tasks (e.g., problem-solving, creative writing, data analysis), among students at different levels of expertise, or between those who have received explicit training in AI literacy and those who have not. Such fine-grained comparisons are more likely to reveal the conditions under which AI supports or undermines higher-order thinking than the coarse disciplinary binary employed in this and many other studies.

This study acknowledges several limitations. The reliance on self-reported HOTS is the most consequential. The positive correlation found here may not replicate if performance-based measures are used, as Gerlich (2025) demonstrated. Future research should combine self-report measures with task-based assessments of critical thinking to determine whether students' perceived engagement in higher-order thinking corresponds to observable cognitive performance. The sample's concentration in Year 1 and Year 2 students means the findings may not generalise to more advanced students whose relationship with AI may be shaped by deeper disciplinary socialisation. Treating AI tools as a homogeneous category is another limitation; a student using ChatGPT for essay generation is engaged in a qualitatively different process from one using Grammarly for proofreading, and future studies should distinguish between these use cases. Finally, the small qualitative sample ($n = 6$), while appropriate for initial thematic exploration, limits the depth and diversity of the insights that can be drawn.

The practical implications are cautiously optimistic. The positive relationship between AI reliance and perceived HOTS suggests that AI tools can be incorporated productively into learning, but this is contingent on students maintaining critical awareness. Educators should design tasks that make evaluative engagement with AI explicit, such as requiring students to compare AI outputs with scholarly sources and justify their acceptance or rejection of AI-generated content (Amdan et al., 2025; Huesca et al., 2024).

6.0 CONCLUSION

This study examined the perceived effects of reliance on AI on higher-order thinking skills among undergraduates at UNIMAS. Contrary to the dominant cognitive offloading narrative, a strong positive association was found between reliance on AI and perceived HOTS engagement, with no significant differences between STEM and non-STEM students. Qualitative data revealed that students' cognitive engagement with AI was shaped less by disciplinary background and more by individual evaluative practices.

These findings do not refute the cognitive offloading hypothesis, but they do qualify it. The relationship between AI use and cognitive engagement is not unidirectional, and the context in which students use AI, their evaluative habits, pedagogical culture, and linguistic needs, may matter more than either the frequency of use or the disciplinary label. The study contributes locally situated evidence from a Malaysian university context that complicates Western-centric assumptions about AI's cognitive risks.

All in all, the study's findings carry implications for pedagogical practice in Malaysian higher education. The positive relationship between reliance on AI and perceived HOTS suggests that blanket restrictions on AI use may be counterproductive. Instead, educators should consider designing learning tasks that make critical engagement with AI an explicit requirement, such as structured comparison exercises in which students evaluate AI-generated responses against scholarly sources and provide reasoned justifications for their decisions. Assessment designs that reward the process of critical evaluation, rather than merely the final product, may help sustain the kind of reflective AI use that the qualitative data describe. At the institutional level, these findings support the integration of AI literacy into curriculum frameworks, equipping students not merely to use AI tools but to understand their limitations and evaluate their outputs critically.

Acknowledgement

We extend our sincere appreciation to all participants for their valuable time, contributions and insights throughout this study.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper

References

Adams, D., & Chuah, K. M. (2022). Artificial intelligence-based tools in research writing: Current trends and future potentials. Artificial intelligence in higher education, 169-184. DOI: https://doi.org/10.1201/9781003184157

Amdan, M. A. B., Janius, N., Saidin, M. S. B., & Kasdiah, A. H. B. (2025). Impact of artificial intelligence in TVET and STEM education among higher learning students in Malaysia. *Journal of Research in Mathematics, Science, and Technology Education, 2*(1), 1–14. DOI: https://doi.org/10.70232/jrmste.v2i1.15

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing writing*, *57*, 100745.

Bowden, A., Fox-Rushby, J., Anyandieka, L., & Wanjau, J. (2002). Methods for pre-testing and piloting survey questions: Illustrations from the KENQOL survey of health-related quality of life. *Health Policy and Planning, 17*(3), 322–330. DOI: https://doi.org/10.1093/heapol/17.3.322

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. DOI: https://doi.org/10.1191/1478088706qp063oa

Çela, E., Fonkam, M., & Potluri, R. M. (2024). Risks of AI-assisted learning on student critical thinking: A case study of Albania. *International Journal of Risk and Contingency Management, 12*(1), 2–15. DOI: https://doi.org/10.4018/IJRCM.350185

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access, 8*, 75264–75278. DOI: https://doi.org/10.1109/ACCESS.2020.2988510

Chuah, K. M., & Sumintono, B. (2024). University student readiness and academic integrity in using ChatGPT and AI tools for assessments. In *Digital assessment in higher education: Navigating and researching challenges and opportunities* (pp. 69-88). Singapore: Springer Nature Singapore. DOI: https://doi.org/10.1007/978-981-97-6136-4_4

Essien, A., Bukoye, O. T., O'Dea, X., & Kremantzis, M. (2024). The influence of AI text generators on critical thinking skills in UK business schools. *Studies in Higher Education, 49*(5), 865–882. DOI: https://doi.org/10.1080/03075079.2024.2316881

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), 1–24. DOI: https://doi.org/10.3390/soc15010006

Guo, Z., Wu, Y., Hartline, J. D., & Hullman, J. (2024). A decision theoretic framework for measuring AI reliance. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)* (pp. 1–17). Association for Computing Machinery. DOI: https://doi.org/10.1145/3630106.3658901

Hasin, H., Jamil, A., Mahat, Z., Sehat, R., & Fauzihardani, E. (2025). The role of AI-assisted learning tools in higher education: Balancing efficiency, critical thinking and academic integrity. *Journal of Social and Development Sciences, 15*(1), 36–46. DOI: https://doi.org/10.22610/jsds.v15i1%28s%29.4462

Huesca, G., Martínez-Treviño, Y., Molina-Espinosa, J. M., Sanromán-Calleros, A. R., Martínez-Román, R., Cendejas-Castro, E. A., & Bustos, R. (2024). Effectiveness of using ChatGPT as a tool to strengthen benefits of the flipped learning strategy. *Education Sciences, 14*(6), 660. DOI: https://doi.org/10.3390/educsci14060660

Khan, N., Sarwar, A., Chen, T. B., & Khan, S. (2022). Connecting digital literacy in higher education to the 21st century workforce. *Knowledge Management & E-Learning*, *14*(1), 46-61. DOI: https://doi.org/10.34105/j.kmel.2022.14.004

Kofahi, M. A., & Husain, A. J. A. (2025). ChatGPT for operating systems: Higher-order thinking in focus. *Journal of Information Technology Education: Research, 24*, 1–23. DOI: https://doi.org/10.28945/5382

Kosar, T., Ostojic, D., Liu, D. Y., & Mernik, M. (2024). Computer science education in ChatGPT era: Experiences from an experiment in a programming course for novice programmers. *Mathematics, 12*(5), 629. DOI: https://doi.org/10.3390/math12050629

Lee, Y., Ahn, H., & Jo, H. (2026). Exploring the drivers and outcomes of generative AI usage: the role of metacognition. *International Journal of Human–Computer Interaction*, *42*(4), 2098-2120. DOI: https://doi.org/10.1080/10447318.2025.2526592

Li, Y., Castulo, N. J., & Xu, X. (2025). Embracing or rejecting AI? A mixed-method study on undergraduate students' perceptions of artificial intelligence at a private university in China. *Frontiers in Education, 10*, 1505856. DOI: https://doi.org/10.3389/feduc.2025.1505856

Lijie, H., Yusoff, S. M., & Marzaini, A. F. (2024). Influence of AI-driven educational tools on critical thinking dispositions among university students in Malaysia: A study of key factors and correlations. *Education and Information Technologies, 30*, 8029–8053. DOI: https://doi.org/10.1007/s10639-024-13150-8

Limna, P., Jakwatanatham, S., Siripipattanakul, S., Kaewpuang, P., & Sriboonruang, P. (2022). A review of artificial intelligence (AI) in education during the digital era. *Advanced Knowledge for Executives, 1*(1), 1–9.

Ministry of Higher Education Malaysia (2025). *Malaysia Higher Education Blueprint (2026-2035)*. Ministry of Higher Education.

Morse, J. M. (2016). *Mixed method design: Principles and procedures*. Routledge.

Obenza, B. N., Salvahan, A., Rios, A. N., Solo, A., Alburo, R. A., & Gabila, R. J. (2024). University students' perception and use of ChatGPT: Generative artificial intelligence (AI) in higher education. *International Journal of Human Computing Studies, 5*(12), 5–18.

Parviz, M. (2024). AI in education: Comparative perspectives from STEM and non-STEM instructors. *Computers and Education Open, 6*, 100190. DOI: https://doi.org/10.1016/j.caeo.2024.100190

Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2017). Dunning–Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic bulletin & review*, *24*(6), 1774-1784. DOI: https://doi.org/10.3758/s13423-017-1242-7

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. DOI: https://doi.org/10.1016/j.tics.2016.07.002

Selvanathan, B., & Narayanan, S. (2024). ChatGPT in higher education Malaysia: An opportunity or threat to the education system? *International Journal of Academic Research in Progressive Education and Development, 13*(3), 965–974. DOI: https://doi.org/10.6007/IJARPED/v13-i3/21455

Shakil, E., & Siddiq, S. (2024). ESL teachers' perceptions about ChatGPT as a threat to analytical writing abilities of ESL learners at graduate level. *Pakistan Languages and Humanities Review, 8*(1), 115–128. DOI: https://doi.org/10.47205/plhr.2024(8-I)10

Valeri, F., Nilsson, P., & Cederqvist, A. (2025). Exploring students' experience of ChatGPT in STEM education. *Computers and Education: Artificial Intelligence, 8*, 100360. DOI: https://doi.org/10.1016/j.caeai.2024.100360

Wilson, D. M., & Narasuman, S. (2020). Investigating teachers' implementation and strategies on higher order thinking skills in school based assessment instruments. *Asian Journal of University Education, 16*(1), 62–75. DOI: https://doi.org/10.24191/ajue.v16i1.8991

Yatani, K., Sramek, Z., & Yang, C. (2024). AI as extraherics: Fostering higher-order thinking skills in human–AI interaction. *arXiv*. DOI: https://doi.org/10.48550/arXiv.2409.09218

Zhou, X., Zhang, J., & Chan, C. (2024). Unveiling students' experiences and perceptions of artificial intelligence usage in higher education. *Journal of University Teaching and Learning Practice, 21*(6), 894–897.