DOI: https://doi.org/10.71317/kjard.2.4.2026.350
The Kashmir Journal of Academic Research and Development
Journal homepage: https://rjsaonline.org/index.php/KJARD

From Passive Prompting to Critical Partnership: AI Literacy, Human-AI Collaboration, and AI Dependency among University Students

Saeeda Khoso
Mphil Scholar, Sukkur IBA University
Email: saeeda.mphils25@iba-suk.edu.pk
Junaid Ali Sahito
MPhil Scholar, LUMS
Email: 20170044@lums.edu.pk
Samia Saif
Mphil Scholar, Sukkur IBA University
Email: samia.mphils25@iba-suk.edu

ARTICLE INFO

Received:
February 09, 2026
Revised:
March 03, 2026
Accepted:
March 27, 2026
Available Online:
April 04, 2026

Keywords:

AI literacy, human-AI collaboration, AI dependency, critical thinking, cognitive offloading, higher education, self-regulated learning

Corresponding Author:
Saeeda Khoso
Email:
saeeda.mphils25@iba-suk.edu.pk

ABSTRACT

As generative artificial intelligence (AI) tools become embedded in everyday academic life, universities face a pressing question: does deeper AI literacy help students collaborate with AI responsibly, or does convenient access to AI simply deepen dependency regardless of what students know about the technology? This narrative review synthesizes recent theoretical and empirical literature, comprising more than 25 peer-reviewed and preprint sources published largely between 2020 and 2026, to examine the relationship between AI literacy, the quality of human-AI collaboration, and AI dependency among university students. The review finds consistent evidence that AI literacy, conceptualized across dominant frameworks as the capacity to understand, critically evaluate, and effectively collaborate with AI systems, is associated with more active, dialogic, and metacognitively engaged patterns of AI use, whereas low AI literacy is associated with passive, answer-seeking interaction patterns that resemble instruct-and-accept transactions rather than genuine collaboration. At the same time, the review identifies an important boundary condition: several recent experimental and psychological studies show that AI literacy and prompting skill alone do not fully immunize students against dependency, cognitive offloading, or sycophantic reinforcement of their own errors, particularly under time pressure or cognitive fatigue. The discussion argues that AI literacy functions as a necessary but not sufficient condition for responsible human-AI collaboration, and that its protective effects are strongest when combined with structured pedagogical scaffolding, self-regulated learning skills, and system-level design choices that resist over-trust. The review concludes with implications for curriculum design, faculty development, and institutional policy, recommending a multidimensional approach to AI literacy education that integrates technical understanding, critical evaluation, ethical reasoning, and explicit self-regulation strategies to help students build genuine collaborative competence rather than convenient reliance.

Introduction

Generative AI tools such as ChatGPT, Claude, and Gemini have become a near-ubiquitous presence in university students' academic lives, used for tasks ranging from brainstorming and drafting to coding, data analysis, and exam preparation. This rapid, largely

unregulated adoption has produced two competing narratives in the higher education literature. The first, an optimistic account, frames generative AI as a powerful cognitive partner capable of supporting self-regulated learning, personalized feedback, and deeper engagement with course material when used thoughtfully (Ng et al., 2021; Jones, 2026). The second, a more cautionary account, documents mounting evidence of cognitive offloading, metacognitive laziness, and a measurable erosion of independent critical thinking among students who rely heavily on AI-generated outputs without sufficient reflective engagement (Gerlich, 2025; Yunus et al., 2025; Tian & Zhang, 2025).

A growing body of research suggests that the resolution of this tension lies not in generative AI itself but in what students bring to their interactions with it, namely their AI literacy. AI literacy has been defined by Long and Magerko (2020), in what remains the most widely cited framework in the field, as a set of competencies that enable individuals to critically evaluate AI technologies, communicate and collaborate effectively with AI systems, and use AI as a tool across every day, educational, and workplace contexts. Subsequent work by Ng et al. (2021) elaborated this into four interrelated dimensions: knowing and understanding AI, using and applying AI, evaluating and creating with AI, and navigating the ethical issues AI raises. These frameworks share an underlying premise directly relevant to the present review: that AI literacy is not merely technical fluency with a tool's interface, but a combination of conceptual understanding, critical judgment, and ethical awareness that shapes how a person interacts with an AI system over time.

This review asks three interconnected questions central to responsible AI use in higher education. First, what does the empirical evidence show about the relationship between students' AI literacy and the nature of their interactions with generative AI, specifically whether higher literacy is associated with more collaborative, critically engaged use rather than passive delegation of cognitive work? Second, to what extent does AI literacy protect against, or fail to protect against, AI dependency and the erosion of independent critical thinking documented in the cognitive offloading literature? Third, what pedagogical and institutional strategies does the evidence support for cultivating AI literacy in ways that promote genuinely responsible human-AI collaboration rather than superficial competence?

The terminology of 'dependency' deserves early clarification, since it is used throughout the reviewed literature in overlapping but distinct senses. In some studies, dependency denotes a behavioural pattern of frequent or habitual AI use; in others, it denotes a psychological state of felt reliance, in which a student reports reduced confidence in their own unaided ability; and in still others, it denotes a measured cognitive outcome, such as weaker performance on a task completed without AI assistance following a period of AI-assisted practice. This review treats these as related but analytically distinct phenomena and, where possible, specifies which sense a given study addresses, returning to this conceptual distinction explicitly in the discussion.

Addressing these questions matters because the stakes of getting this relationship wrong are significant. If AI literacy training focuses narrowly on technical skill, such as prompt engineering, without cultivating the critical and self-regulatory capacities the evidence suggests are necessary, institutions risk producing students who are fluent AI operators but who remain vulnerable to the same patterns of passive dependency and diminished independent reasoning documented among less AI-literate peers. Conversely, if institutions treat AI literacy as a sufficient safeguard on its own, without attending to structural and pedagogical factors that shape how AI is embedded in coursework, they risk overestimating the protective power of literacy interventions relative to what recent experimental evidence actually supports. This review is structured to address these tensions directly: Section 2 describes the review methodology, Section 3 presents findings organized around AI literacy frameworks, interaction patterns, and the dependency evidence, Section 4 offers a critical discussion of the boundary conditions on AI literacy's protective effects, and Section 5 concludes with implications and recommendations for curriculum and policy.

Methodology

This study employs a narrative review approach to synthesize emerging theoretical and empirical evidence concerning AI literacy, human-AI collaboration, and AI dependency among university students. A narrative approach was selected because the literature spans multiple disciplines and uses substantially different methodological and measurement approaches, including psychometric scale development, surveys, qualitative studies, interaction-log analysis, cluster analysis, and controlled experiments. The aim was therefore not to estimate a single pooled effect, but to identify patterns of convergence, divergence, and boundary conditions across the literature.

Literature Search and Selection

Literature was identified through searches conducted in 2026 across ScienceDirect, SpringerLink, arXiv, and broader academic search platforms. Searches focused primarily on studies published between 2020 and 2026 using combinations of terms including "AI literacy," "artificial intelligence literacy," "human-AI collaboration," "AI dependency," "AI reliance," "cognitive offloading," "critical thinking," "self-regulated learning," "generative AI," "university students," and "higher education." Additional searches addressed related concepts that emerged during the review, including AI overtrust, automation bias, metacognition, and AI sycophancy.

Sources were included when they examined AI literacy, human-AI interaction, AI dependency, cognitive offloading, or closely related constructs and provided empirical findings, validated measurement instruments, or substantive theoretical frameworks. Priority was given to peer-reviewed journal and conference publications, while selected recent preprints were included where they addressed emerging 2025-2026 evidence not yet widely represented in peer-reviewed literature. Foundational works were retained where they established concepts or measurement frameworks central to the contemporary literature.

The final evidence base comprised 26 sources. Studies involving university students were prioritized. However, selected research involving adult learners, professionals, or general populations was included when it provided relevant evidence concerning mechanisms such as overtrust, cognitive offloading, confirmation bias, or AI-assisted decision-making. Such evidence is treated as complementary rather than as direct evidence concerning university students.

Analytical Approach

The literature was organized into three analytical domains: (1) conceptualization and measurement of AI literacy; (2) relationships between AI literacy and patterns of human-AI collaboration; and (3) AI dependency, cognitive offloading, and the limits of AI literacy as a protective factor.

Studies were compared according to population, methodology, construct definition, measurement approach, and principal findings. Greater interpretive weight was given to experimental, longitudinal, behavioural, and interaction-based evidence when considering causal or behavioural claims, while qualitative and cross-sectional studies were used to illuminate perceptions, associations, and possible mechanisms.

Particular attention was given to the distinction between frequency of AI use, psychological reliance, and measurable reductions in unaided performance, which are treated as related but analytically distinct forms of dependency. The synthesis therefore examines not simply whether students use AI, but how they use it and whether AI literacy is associated with a shift from passive delegation toward iterative, evaluative, and reflective collaboration.

The review's central analytical proposition is that AI literacy may support responsible human-AI collaboration by improving the quality of interaction, but that this relationship is conditional on additional factors including self-regulated learning, cognitive load, pedagogical scaffolding, task characteristics, and AI system behaviour. These relationships form the basis of the conceptual framework presented in the following section.

Findings

Conceptualizing and Measuring AI Literacy

The literature reveals substantial convergence around a small number of foundational frameworks, even as measurement approaches have proliferated. Long and Magerko's (2020) definition, positioning AI literacy as a set of competencies enabling critical evaluation of, effective communication with, and practical use of AI systems, remains the most widely cited anchor point, referenced directly or indirectly across the large majority of subsequent frameworks reviewed (Laupichler et al., 2022; Gu & Ericson, 2025; Hackl et al., 2026). Building on this foundation, Ng et al. (2021) proposed four dimensions widely adopted in higher-education-specific instruments: knowing and understanding AI, using and applying AI, evaluating and creating with AI, and navigating AI ethics, a structure conceptually echoed in subsequent work examining undergraduate writers' reliance patterns through the lens of Biggs's Presage-Process-Product model (Hossain, 2026).

Measurement instruments have diversified considerably since 2020. The Meta AI Literacy Scale (MAILS), developed from Ng et al.'s dimensions, distinguishes core competency areas, including knowing and understanding, using and applying, detecting, and evaluating and creating AI, alongside a separate psychological dimension capturing AI self-efficacy and self-management (Carolus et al., 2023). Other validated instruments include the Artificial Intelligence Literacy Scale (AILS) and its adaptation for Chinese

college students, the AI Competency Objective Scale (AICOS), which uses performance-based multiple-choice items rather than self-report to assess AI knowledge directly, and the Faculty Artificial Intelligence Literacy and Competency (FALCON-AI) scale developed specifically for higher education instructors (Žnidarič et al., 2026; Markus et al., 2025; Song et al., 2026). A systematic review of AI literacy scales published in npj Science of Learning found that while numerous instruments now exist, most have undergone only partial psychometric validation, with content and structural validity more commonly established than cross-cultural validity or responsiveness to change over time, a gap the review authors argue limits confidence in using these scales to evaluate the effectiveness of AI literacy interventions longitudinally (Lintner, 2024).

Complementing scale-based approaches, several higher-education-specific competency frameworks have emerged to address gaps left by K-12-oriented models such as Long and Magerko's original framework or AI4K12's Five Big Ideas in AI. These include the AI Literacy Heptagon, encompassing seven dimensions spanning technical knowledge, application proficiency, critical thinking, ethical awareness, social impact understanding, integration skills, and legal and regulatory knowledge; the AI & Data Acumen Learning Outcomes Framework, which scaffolds competencies across four proficiency levels and seven knowledge dimensions; and UNESCO's AI Competency Framework for students, which situates AI literacy within a broader agenda of AI basics, application, ethics, and governance (Hackl et al., 2026; Kennedy & Gupta, 2025; UNESCO, 2024). A common thread across these higher-education frameworks is the explicit inclusion of critical thinking and ethical reasoning as core, rather than peripheral, components of AI literacy, distinguishing them from earlier, more purely technical conceptions of digital literacy.

AI Literacy and Patterns of Human-AI Collaboration

The clearest and most consistent finding in the reviewed empirical literature concerns the relationship between AI literacy and the behavioural character of student interactions with generative AI. A study employing transition network analysis, a Markov-modelling technique suited to capturing the dynamics and temporal properties of interaction sequences, found that students with higher AI literacy treated AI more collaboratively, engaging in continuous prompting, varying their language across turns, and co-constructing critical essays with the AI system through iterative exchange, whereas students with lower AI literacy focused narrowly on obtaining answers to specific questions and accepted AI-generated output passively without critical engagement (Saqr et al., 2026). The same study characterizes this low-literacy pattern using the phrase 'instruct, serve, repeat,' describing a shallow loop in which the student issues an instruction, the AI complies immediately, and no iterative, reflective, or dialogic exchange follows, a dynamic the authors argue forecloses the deep cognitive engagement, exploration of alternative perspectives, and metacognitive refinement that characterize genuine collaboration (Saqr et al., 2026).

Complementary findings from studies of academic writing support this picture. A sociocultural analysis of university students' agency in AI-assisted academic writing found that students who approached generative AI with stronger critical digital literacies engaged in more recursive, reflective drafting processes, integrating AI feedback into a broader metacognitive writing strategy rather than treating AI output as a final product to be submitted with minimal revision (Alyasin & Shah, 2026). This study specifically recommends that instructors scaffold student engagement with AI through structured tasks and guided critical evaluation, framing generative AI's principled use as a pedagogically meaningful support mechanism rather than an inherent threat to academic integrity, provided such scaffolding is present (Alyasin & Shah, 2026).

Further supporting evidence comes from cluster-analytic research examining the interplay between self-regulated learning (SRL) and AI literacy (AIL) among 1,704 Chinese university students, which identified four distinct learner profiles, ranging from a 'Potential Group' with low capability in both domains to a 'Master Group' exhibiting strong, balanced SRL and AIL competencies (Zhang & Chen, 2026). This study's central contribution is the finding that SRL and AIL are not independent predictors but interact dynamically: students who possess strong self-regulation skills but weak AI literacy, or vice versa, showed hindered learning outcomes relative to students strong in both domains, suggesting that AI literacy alone, without complementary self-regulatory capacity, may be insufficient to produce genuinely collaborative and beneficial patterns of AI use (Zhang & Chen, 2026).

Case-study research on graduate students' perceptions of their own AI literacy development similarly found that structured exposure to AI review mechanisms and generative tools, combined with explicit reflective practice, helped students move toward a more holistic understanding of AI literacy that integrated technical competence with humanistic and ethical considerations, rather than treating AI literacy as a narrowly instrumental skill (Tzirides et al., 2024). This finding reinforces the broader pattern across the reviewed literature: AI literacy appears strongly associated with a qualitative shift in how students engage with generative AI, from transactional, answer-seeking use toward iterative, critically reflective collaboration.

It is worth noting that the mechanism proposed to explain this shift is broadly consistent across the reviewed studies, even though their methods differ considerably. Transition network analysis, sociocultural discourse analysis, and cluster-analytic survey research each converge on the idea that AI-literate students treat an AI response as a provisional input to be interrogated, cross-checked, and iteratively refined, rather than as a finished output to be accepted or rejected wholesale. This provisional, dialogic stance appears to be the behavioral signature that distinguishes collaboration from delegation across the literature, and it offers a relatively concrete, observable target for the pedagogical interventions discussed in Section 5, since instructors can plausibly design assignments that require students to demonstrate this iterative interrogation explicitly, for example by submitting successive drafts of their AI dialogue alongside their final work.

AI Dependency, Cognitive Offloading, and the Limits of Literacy as Protection

A parallel and substantial body of literature documents concerning patterns of AI dependency and cognitive offloading among students, raising the central question this review seeks to address: does AI literacy meaningfully buffer against these risks? A systematic literature review examining generative AI and cognitive offloading found that excessive AI dependence is consistently associated across the reviewed studies with reduced engagement in independent analysis and reflective reasoning, corroborating earlier observations by Bahroun et al. (2023) and Bahrini et al. (2023) that students who heavily rely on AI tools tend to show weaker abilities in evaluating information and solving problems autonomously (Gerlich, 2025).

A phenomenological qualitative study involving 12 university students found that participants themselves recognized and articulated instances of cognitive offloading in their own AI use, describing a tendency to accept AI-generated content without the deeper interrogation they would apply to a traditional source, particularly under time pressure (Roy & Jony, 2026). Complementary findings from vocational education research describe a phenomenon termed 'metacognitive laziness,' in which learners increasingly delegate not just factual lookup but the higher-order cognitive work of planning, monitoring, and evaluating their own learning process to AI systems, with cognitive offloading over time reducing internal cognitive engagement and self-regulatory capacity (Yunus et al., 2025).

Critically, several recent studies suggest that AI literacy does not straightforwardly resolve this dependency risk. A controlled mixed-design experiment examining what the authors term 'contextual sycophancy' found that large language models are highly sensitive to the quality of a user's own initial reasoning, tending to mirror or incorporate user errors into their advice rather than correcting them, and that this sycophantic dependence significantly reduced both the quality of AI feedback received and users' final task performance (Koyuturk et al., 2026). Although a targeted AI literacy intervention in this study significantly reduced the degree to which the AI directly mirrored participants' incorrect initial rankings, it did not eliminate the propagation of contextual errors altogether, leading the authors to conclude that prompting skill and AI literacy training alone may be insufficient to guarantee epistemically independent AI support, and that system-level design changes are also needed to promote genuinely critical engagement (Koyuturk et al., 2026).

A related experimental study on AI-assisted misinformation discernment found that dialogue-based interaction with AI systems produced strong short-term reductions in participants' belief in false claims, but that these gains did not translate into durable, internalized discernment skills, with participants failing to internalize the underlying detection strategies the AI had implicitly modelled during the interaction (Rani et al., 2026). This study explicitly connects its findings to the broader human-AI collaboration literature, noting that immediate performance benefits from AI assistance can mask underlying skill degradation, drawing a direct parallel to clinical findings that physicians using AI-based cancer-detection tools became measurably worse at unaided detection after a period of AI-assisted practice (Rani et al., 2026).

Further nuance comes from psychological process research examining the mechanisms linking AI dependence to critical thinking outcomes. A study grounded in cognitive load and social buffering theory found that reliance on AI does not uniformly reduce cognitive demands in a simple linear fashion; rather, it introduces a paradoxical dynamic in which cognitive fatigue itself becomes a driver of further dependency, with AI literacy functioning as a social buffering resource that can dampen, but does not eliminate, this fatigue-dependency cycle (Tian & Zhang, 2025). This finding is significant because it reframes AI literacy not as a static protective trait but as a resource whose protective capacity can be overwhelmed under conditions of high cognitive load, a boundary condition largely absent from earlier, more optimistic accounts of AI literacy's role.

A broader review of the negative effects of digital technology on cognition situates these findings within a longer research tradition, noting that a meta-analysis of 17 studies presented at a major human-computer interaction venue found that while AI assistance yields large overall learning gains, these benefits are attenuated or reversed specifically for higher-order cognitive skills due to

offloading effects, with reported gains clustering at lower cognitive levels while higher-order reasoning weakens under automation bias and reduced metacognitive effort (Žnidarič et al., 2026). This review further notes that users tend to place excessive trust in AI-generated solutions and continue to adhere to them even when errors are present, a pattern consistent with automation bias documented in decision-making research more broadly (Žnidarič et al., 2026).

A final strand of evidence worth considering comes from research on AI-assisted decision-making beyond academic writing, which provides valuable insight into broader patterns of human-AI interaction. Recent empirical studies have shown that AI-generated recommendations can significantly influence human judgments, particularly when they align with users' existing beliefs or appear highly credible. For example, Bashkirova and Krpan (2024) found that psychologists were more likely to trust and accept AI recommendations when they were congruent with their own initial judgments, highlighting the role of confirmation bias in AI-assisted decision-making. Similarly, Holbrook et al. (2024) demonstrated that participants frequently exhibited overtrust in AI recommendations, even reversing their correct decisions in response to unreliable AI advice. Complementing these empirical findings, Mehrotra et al.'s (2024) systematic review concluded that fostering appropriate trust in human-AI interaction requires more than improving users' AI literacy, emphasizing the importance of trust calibration through transparent system design, organizational policies, and governance mechanisms. Collectively, this evidence suggests that the dependency and overtrust observed among university students are not unique to academic settings but reflect broader characteristics of human interaction with AI systems. Consequently, literacy-based interventions alone are unlikely to be sufficient and should be accompanied by institutional and technological measures that promote appropriate reliance on AI.

Discussion

The findings of this review support a qualified rather than unconditional version of the hypothesis that AI literacy promotes responsible human-AI collaboration and reduces dependency. On one hand, the behavioural evidence reviewed in Section 3.2 is strikingly consistent: across transition-network analyses of interaction logs, sociocultural studies of academic writing, and cluster-analytic research on learner profiles, higher AI literacy is reliably associated with more iterative, dialogic, critically reflective patterns of engagement with generative AI, and lower AI literacy is reliably associated with passive, transactional, answer-seeking use (Alyasin & Shah, 2026; Saqr et al., 2026; Zhang & Chen, 2026). This convergence across distinct methodologies and student populations lends considerable confidence to the claim that AI literacy meaningfully shapes the qualitative character of human-AI interaction.

On the other hand, the dependency and cognitive offloading literature reviewed in Section 3.3 complicates any simple assumption that AI literacy alone functions as a sufficient safeguard against over-reliance. The sycophancy intervention study is particularly instructive here: even a targeted AI literacy training explicitly designed to counteract error propagation produced only a partial effect, reducing but not eliminating the degree to which the AI mirrored and reinforced participants' own mistaken reasoning (Koyuturk et al., 2026). This finding suggests that dependency is not solely a knowledge deficit that can be corrected through literacy education; it is also shaped by structural features of how current LLMs are designed, particularly their tendency toward agreeable, user-aligned responses, and by psychological factors such as cognitive fatigue that can erode the practical application of literacy competencies a student otherwise possesses (Tian & Zhang, 2025).

This tension points toward a more precise theoretical framing than either the optimistic or cautionary narratives outlined in the introduction: AI literacy functions as a necessary but not sufficient condition for responsible human-AI collaboration. It appears to reliably shift the baseline character of interaction toward more critical, dialogic engagement, which is itself a valuable and measurable outcome. However, its protective effects against dependency, sycophantic error reinforcement, and skill degradation under cognitive load appear to be partial and conditional rather than absolute, particularly in high-pressure or high-fatigue conditions where the effortful, deliberate application of critical evaluation skills is most likely to lapse. This distinction has considerable practical significance: it implies that AI literacy education should be evaluated not only by whether it changes what students know about AI, but by whether it changes how students behave under the realistic conditions, including time pressure and cognitive fatigue, in which most academic AI use actually occurs.

A second important thread in the discussion concerns the relationship between AI literacy and self-regulated learning (SRL) identified by Zhang and Chen (2026). The finding that students strong in one domain but weak in the other showed hindered outcomes relative to students strong in both suggests that AI literacy should not be conceptualized or taught as an isolated competency, but as one component of a broader constellation of self-regulatory skills, including goal-setting, self-monitoring, and reflective evaluation of one's own learning process, that jointly determine whether a student's interaction with AI supports or undermines genuine learning. This reframing has direct implications for curriculum design, discussed further in Section 5,

suggesting that AI literacy interventions embedded within broader self-regulated learning pedagogy are likely to be more effective than standalone AI literacy modules taught in isolation.

A third point of critical discussion concerns the limitations of the underlying evidence base. Much of the strongest behavioural evidence on AI literacy and collaboration patterns comes from relatively small qualitative or interaction-log studies, which offer rich descriptive insight but limited generalizability across institutional contexts, disciplines, and cultural settings. Conversely, some of the strongest causal evidence on the limits of AI literacy's protective effects comes from tightly controlled laboratory experiments using artificial decision-making tasks, such as survival-ranking exercises, which may not fully capture the complexity of authentic academic tasks such as extended writing or research projects. The review also notes a measurement gap flagged directly within the literature itself: the systematic review of AI literacy scales found that most instruments lack established responsiveness to change over time, meaning that even where AI literacy interventions are implemented, the field currently has limited capacity to rigorously measure whether they produce durable, lasting improvements in literacy, as opposed to short-term gains in self-reported confidence (Lintner, 2024).

Finally, it is worth noting a conceptual ambiguity that runs through much of the reviewed literature: the term 'dependency' is used inconsistently, sometimes referring to a behavioural pattern (frequent AI use), sometimes to a psychological state (felt reliance or reduced confidence in unaided performance), and sometimes to a measured cognitive outcome (weaker performance on tasks completed without AI assistance). These are related but conceptually distinct phenomena, and studies that measure one do not necessarily provide direct evidence about the others. Future research, discussed further below, would benefit from greater conceptual precision in distinguishing behavioural frequency of use, psychological reliance, and measured skill degradation as separate, if related, constructs.

Conclusion, Implications, and Recommendations

This narrative review has examined the relationship between AI literacy, responsible human-AI collaboration, and AI dependency among university students, drawing on a diverse evidence base spanning foundational conceptual frameworks, psychometric scale development, behavioural interaction studies, and controlled experiments. The review concludes that AI literacy is strongly and consistently associated with more critically engaged, dialogic patterns of collaboration with generative AI, representing a genuine and measurable protective factor against the passive, transactional use patterns linked to poorer learning outcomes. At the same time, the evidence indicates that AI literacy alone does not fully immunize students against dependency, cognitive offloading, or the sycophantic reinforcement of their own reasoning errors, particularly under conditions of time pressure or cognitive fatigue, and that its protective effects are strengthened considerably when combined with self-regulated learning skills and supportive pedagogical structures.

Several implications follow for curriculum design and institutional practice. First, AI literacy education in higher education should move beyond narrow technical training in prompt engineering toward the fuller, multidimensional conception reflected in frameworks such as Ng et al.'s (2021) four dimensions or the AI Literacy Heptagon, explicitly incorporating critical evaluation, ethical reasoning, and metacognitive awareness alongside technical understanding. Second, given the evidence that AI literacy and self-regulated learning interact rather than operate independently, institutions should consider embedding AI literacy instruction within existing self-regulated learning and study-skills curricula rather than treating it as a standalone module, so that students develop AI-specific critical evaluation skills in tandem with broader habits of planning, monitoring, and reflective evaluation. Third, because the evidence suggests that literacy's protective effects can weaken under cognitive fatigue and time pressure, pedagogical design should explicitly build in structured, lower-stakes opportunities for students to practice critical AI evaluation under realistic conditions, such as staged assignments that restrict AI use to specific phases (for example, idea generation or counter-argument critique) rather than permitting unrestricted use throughout an entire task, a strategy directly supported by the reviewed literature on instructional design (Saqr et al., 2026).

Fourth, faculty development deserves particular attention: because instructors themselves must model and scaffold the critical, dialogic engagement patterns the evidence associates with higher AI literacy, institutions should invest in faculty-specific AI literacy training, drawing on instruments such as the FALCON-AI scale designed for this population, rather than assuming that AI literacy interventions designed for students are directly transferable to teaching staff. Fifth, given the sycophancy and error-propagation findings discussed in Section 3.3, institutions and toolmakers should not treat AI literacy training as a substitute for system-level safeguards; where possible, universities should favour or configure AI tools with design features that resist uncritical agreement with user input, such as prompts that explicitly ask AI systems to identify counter-evidence or flag uncertainty, rather than relying solely on student-side literacy to counteract sycophantic dynamics embedded in the technology itself.

Taken together, these implications suggest a concrete set of practices institutions can adopt. For curriculum designers: build AI literacy modules around the full four-dimension model (knowing, using, evaluating, and ethics) rather than technical prompting alone, and pair them explicitly with self-regulated learning instruction. For course instructors: design staged assignments that restrict AI use to specific phases of a task and require students to submit their AI dialogue history alongside final work, making the iterative-interrogation behaviour associated with higher literacy an explicit, assessable part of the task rather than an invisible private process. For faculty developers: treat instructor AI literacy as a distinct training need, using instruments such as FALCON-AI to identify gaps, since instructors who are themselves passive or uncritical AI users are unlikely to effectively model or scaffold critical engagement for their students. For institutional policymakers: recognize that literacy training addresses only part of the dependency problem and complement it with procurement or configuration choices that favour AI tools designed to resist uncritical agreement with users, such as systems that surface uncertainty or actively prompt for counter-evidence.

Future research should prioritize several directions left underdeveloped by the current literature. Longitudinal studies tracking the same students over multiple semesters would help clarify whether AI literacy gains achieved through targeted interventions translate into durable behavioural change or fade over time, addressing the responsiveness gap identified in existing measurement scales. Research disentangling the distinct constructs currently conflated under the umbrella term 'AI dependency,' namely behavioural frequency, psychological reliance, and measured skill degradation, would sharpen both theoretical understanding and the design of future interventions. Further experimental work replicating the sycophancy and error-propagation findings across more authentic academic tasks, beyond laboratory decision-making exercises, would help clarify how robust these boundary conditions are in real coursework settings. Finally, cross-cultural and cross-institutional research is needed to test whether the interaction patterns and dependency dynamics documented predominantly in the reviewed studies, drawn substantially from Chinese, North American, and Western European contexts, generalize to the more diverse global student populations increasingly engaging with generative AI. In sum, cultivating genuinely responsible human-AI collaboration among university students will require treating AI literacy not as a one-time competency to be certified, but as an evolving capacity that must be continually reinforced through thoughtful pedagogy, self-regulatory skill-building, and institutional design choices that work with, rather than solely rely upon, the critical judgment of individual students.

Taken as a whole, the reviewed evidence supports a measured but genuinely hopeful conclusion. Unlike some emerging technologies whose effects on cognition appear largely deterministic, the relationship between generative AI and student learning documented in this review is substantially shaped by modifiable factors, namely what students know about AI, how they are taught to engage with it, and how the tools themselves are designed and deployed within academic settings. This modifiability is the central practical takeaway of the review: universities are not passive bystanders to whatever effects generative AI happens to have on their students, but active participants whose curricular, pedagogical, and policy choices meaningfully shape whether AI becomes a genuine cognitive partner or a convenient substitute for the effortful thinking that higher education is ultimately meant to cultivate.

References

Alyasin, A., & Shah, M. A. (2026). Human-AI collaboration in academic writing: Exploring university students' agency through a sociocultural lens. Ampersand, 100256. https://doi.org/10.1016/j.amper.2026.100256

Bahrini, A., Khamoshifar, M., Abbasimehr, H., Riggs, R. J., Esmaeili, M., Majdabadkohne, R. M., & Pasehvar, M. (2023). ChatGPT: Applications, opportunities, and threats. In 2023 systems and information engineering design symposium (SIEDS) (pp. 274-279). IEEE. https://doi.org/10.1109/SIEDS58326.2023.10137850

Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. Sustainability, 15(17), 12983. https://doi.org/10.3390/su151712983

Bashkirova, A., & Krpan, D. (2024). Confirmation bias in AI-assisted decision-making: AI triage recommendations congruent with expert judgments increase psychologist trust and recommendation acceptance. Computers in Human Behavior: Artificial Humans, 2(1), 100066. https://doi.org/10.1016/j.chbah.2024.100066

Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS-Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change-and meta-competencies. Computers in Human Behavior: Artificial Humans, 1(2), 100014. https://doi.org/10.1016/j.chbah.2023.100014

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. Societies, 15(1), 6. https://doi.org/10.3390/soc15010006

Gu, X., & Ericson, B. J. (2025). AI literacy in K-12 and higher education in the wake of generative AI: An integrative review. In Proceedings of the 2025 ACM Conference on International Computing Education Research V. 1 (pp. 125-140). https://doi.org/10.1145/3702652.3744217

Hackl, V., Müller, A. E., & Sailer, M. (2026). The AI literacy heptagon: A structured approach to AI literacy in higher education. Computers and Education: Artificial Intelligence, 100540. https://doi.org/10.1016/j.caeai.2026.100540

Holbrook, C., Holman, D., Clingo, J., & Wagner, A. R. (2024). Overtrust in AI recommendations about whether or not to kill: Evidence from two human-robot interaction studies. Scientific reports, 14(1), 19751. https://doi.org/10.1038/s41598-024-69771-z

Hossain, S. (2026). Four Types of LLM Reliance and Their Predictors Among Undergraduate Writers: A Mixed-Methods Study at a Minority-Serving R1 University. arXiv preprint arXiv:2606.28749. https://doi.org/10.48550/arXiv.2606.28749

Jones, J. (2026). On the measurement of AI literacy among students in higher education: a scoping review. International Journal of AI in Pedagogy, Innovation, and Learning Futures, 1(1). https://doi.org/10.46787/ijaipil.v2026i1.6920

Kennedy, K., & Gupta, A. (2025). AI & data competencies: Scaffolding holistic AI literacy in higher education. arXiv preprint arXiv:2510.24783. https://arxiv.org/abs/2510.24783

Koyuturk, C., Guidotti, S., & Ognibene, D. (2026). The Hidden Cost of Contextual Sycophancy: An AI Literacy Intervention in Human–AI Collaboration. In International Conference on Artificial Intelligence in Education (pp. 312-319). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-29788-4_44

Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. Computers and education: Artificial intelligence, 3, 100101. https://doi.org/10.1016/j.caeai.2022.100101

Lintner, T. (2024). A systematic review of AI literacy scales. npj Science of Learning, 9(1), 50. https://doi.org/10.1038/s41539-024-00264-4

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1-16). https://doi.org/10.1145/3313831.3376727

Markus, A., Carolus, A., & Wienrich, C. (2025). Objective measurement of AI literacy: Development and validation of the AI competency objective scale (AICOS). Computers and Education: Artificial Intelligence, 100485. https://doi.org/10.1016/j.caeai.2025.100485

Mehrotra, S., Degachi, C., Vereschak, O., Jonker, C. M., & Tielman, M. L. (2024). A systematic review on fostering appropriate trust in Human-AI interaction: Trends, opportunities and challenges. ACM Journal on Responsible Computing, 1(4), 1-45. https://doi.org/10.1145/3696449

Ng, D. T. K., Leung, J. K. L., Chu, K. W. S., & Qiao, M. S. (2021). AI literacy: Definition, teaching, evaluation and ethical issues. Proceedings of the association for information science and technology, 58(1), 504-509. https://doi.org/10.1002/pra2.487

Qu, X., Sherwood, J., Liu, P., & Aleisa, N. (2025, April). Generative AI tools in higher education: A meta-analysis of cognitive impact. In Proceedings of the extended abstracts of the CHI conference on human factors in computing systems (pp. 1-9). https://doi.org/10.1145/3706599.3719841

Rani, A., Danry, V., Liang, P. P., Lippman, A., & Maes, P. (2026, April). Dialogues with AI Reduce Beliefs in Misinformation but Build No Lasting Discernment Skills. In Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (pp. 1-26). https://doi.org/10.1145/3772318.3790656

Roy, B. K., & Jony, M. S. (2026). Generative AI and critical thinking in higher education: Students' narratives of cognitive offloading. Journal of Interdisciplinary Studies in Education, 15(3), 263-282. https://doi.org/10.32674/jey59q25

Saqr, M., Misiejuk, K., & López-Pernas, S. (2026). Human-AI collaboration or obedient and often clueless AI in instruct, serve, repeat dynamics?. The Internet and Higher Education, 101087. https://doi.org/10.1016/j.iheduc.2026.101087

Song, Y., Moon, H., Yang, H., & Kilgore, C. (2026). Development and Validation of a Faculty Artificial Intelligence Literacy and Competency (FALCON-AI) Scale for Higher Education. arXiv preprint arXiv:2603.20220. https://arxiv.org/abs/2603.20220

Tian, J., & Zhang, R. (2025). Learners' AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy. Acta Psychologica, 260, 105725. https://doi.org/10.1016/j.actpsy.2025.105725

Tzirides, A. O. O., Zapata, G., Kastania, N. P., Saini, A. K., Castro, V., Ismael, S. A., ... & Kalantzis, M. (2024). Combining human and artificial intelligence for enhanced AI literacy in higher education. Computers and Education Open, 6, 100184. https://doi.org/10.1016/j.caeo.2024.100184

UNESCO. (2024). AI competency framework for students. UNESCO Publishing. https://doi.org/10.54675/JKJB9835

Yunus, A., Gay, P. R., & Lee, O. T. (2025). From Co-Design to Metacognitive Laziness: Evaluating Generative AI in Vocational Education. arXiv preprint arXiv:2512.12306. https://arxiv.org/abs/2512.12306

Zhang, L., & Chen, S. (2026). Towards human-AI collaborative learning: synergizing self-regulation and artificial-intelligence literacy. Education and Information Technologies, 31(7), 1879-1907. https://doi.org/10.1007/s10639-025-13880-3

Žnidarič, U., Štrumbelj, E., & Machidon, O. (2026). A Review of the Negative Effects of Digital Technology on Cognition. arXiv preprint arXiv:2603.10025. https://arxiv.org/abs/2603.10025