

BEYOND CHATGPT: EXAMINING THE IMPACT OF GENERATIVE ARTIFICIAL INTELLIGENCE LITERACY ON HIGHER ORDER THINKING SKILLS THROUGH SELF-REGULATED LEARNING AMONG UNIVERSITY STUDENTS**MELAMPAUI CHATGPT: MENGAJI PENGARUH LITERASI KECERDASAN BUATAN GENERATIF TERHADAP KETERAMPILAN BERPIKIR TINGKAT TINGGI MELALUI PEMBELAJARAN YANG DIATUR SENDIRI PADA MAHASISWA DI UNIVERSITAS****Danny Yonathan**

Sekolah Tinggi Teologi Berita Hidup, Central Java, Indonesia
yonathandanny1@gmail.com

Corresponding Author*ABSTRACT**

University classrooms did not so much adopt generative artificial intelligence as get overtaken by it, and the pedagogical thinking has been trying to catch up ever since. This narrative review looks at how the different dimensions of generative AI (GenAI) literacy relate to university students' Higher Order Thinking Skills (HOTS), and more specifically at the part Self-Regulated Learning (SRL) plays in deciding whether a chatbot ends up sharpening a student's thinking or quietly doing that thinking for them. Literature was drawn mainly from Scopus- and Web of Science-indexed journals, several arXiv preprints, and major international survey reports published between roughly 2023 and 2026, then organised around four anchors: AI literacy frameworks, Bloom's Revised Taxonomy, Zimmerman's cyclical SRL model, and the empirical record on cognitive offloading and academic integrity. Read together, this literature treats GenAI literacy as a multidimensional construct covering technical, critical-evaluative, ethical, and metacognitive competencies, and it treats the GenAI-HOTS relationship as conditional rather than direct: the same tool produces very different outcomes depending on whether its use is structured and metacognitively guided or left unstructured. SRL is what tips the balance. It is the variable that turns GenAI into either a scaffold for analysis, evaluation, and creation, or a substitute that lets those skills quietly atrophy. The review closes with implications for curriculum design, arguing that GenAI literacy and SRL need to be taught together, deliberately, rather than assumed to develop on their own from mere exposure, and it sketches a research agenda addressing the field's continuing shortage of longitudinal and non-Western evidence.

Keywords: generative artificial intelligence literacy; higher order thinking skills; self-regulated learning; higher education; cognitive offloading; curriculum design

ABSTRAK

Ruang-ruang pembelajaran di perguruan tinggi tidak sekadar mengadopsi kecerdasan buatan generatif (Generative Artificial Intelligence/GenAI), tetapi justru mengalami transformasi yang sangat cepat akibat kehadirannya, sementara pemikiran pedagogis masih terus berupaya menyesuaikan diri. Artikel tinjauan naratif ini mengkaji bagaimana berbagai dimensi literasi kecerdasan buatan generatif berhubungan dengan Higher Order Thinking Skills (HOTS) mahasiswa perguruan tinggi, serta secara khusus menelaah peran Self-Regulated Learning (SRL) dalam menentukan apakah penggunaan chatbot mampu mempertajam kemampuan berpikir mahasiswa atau justru secara perlahan mengambil alih proses berpikir tersebut. Literatur yang dianalisis terutama berasal dari jurnal yang terindeks Scopus dan Web of Science, beberapa preprint arXiv, serta laporan survei internasional utama yang diterbitkan selama periode sekitar tahun 2023–2026. Kajian ini kemudian disusun berdasarkan empat landasan utama, yaitu kerangka literasi kecerdasan buatan, Taksonomi Bloom Revisi, model siklus Self-Regulated Learning dari Zimmerman, serta temuan empiris mengenai cognitive offloading dan integritas akademik. Secara keseluruhan, literatur menunjukkan bahwa literasi GenAI merupakan konstruk multidimensional yang mencakup kompetensi teknis, kemampuan berpikir kritis dan evaluatif, kompetensi etis, serta kompetensi metakognitif. Selain itu, hubungan antara literasi GenAI dan HOTS bersifat kondisional, bukan hubungan

langsung. Alat yang sama dapat menghasilkan dampak yang sangat berbeda bergantung pada apakah penggunaannya dirancang secara terstruktur dan dibimbing melalui strategi metakognitif atau digunakan tanpa arahan yang jelas. Dalam konteks tersebut, Self-Regulated Learning (SRL) menjadi faktor penentu yang mengarahkan keseimbangan. SRL menjadikan GenAI sebagai scaffold yang mendukung kemampuan analisis, evaluasi, dan kreasi, atau sebaliknya sebagai pengganti yang menyebabkan kemampuan-kemampuan tersebut secara perlahan mengalami penurunan. Artikel ini ditutup dengan implikasi bagi pengembangan kurikulum, yaitu bahwa literasi GenAI dan Self-Regulated Learning perlu diajarkan secara terpadu dan disengaja, bukan diasumsikan akan berkembang secara otomatis hanya melalui paparan terhadap teknologi. Selain itu, artikel ini mengusulkan agenda penelitian yang berfokus pada keterbatasan bukti empiris saat ini, khususnya masih minimnya penelitian longitudinal serta penelitian yang dilakukan di luar konteks negara-negara Barat.

Kata Kunci: literasi kecerdasan buatan generatif; keterampilan berpikir tingkat tinggi; pembelajaran yang diatur sendiri; pendidikan tinggi; cognitive offloading; desain kurikulum.

1. INTRODUCTION

It took less than four years for ChatGPT and the generative AI (GenAI) systems that followed it to go from novelty to near-default academic infrastructure on university campuses worldwide. The scale of that shift shows up consistently across several independent survey lines. The Higher Education Policy Institute's annual UK student survey found that the share of students reporting any use of AI tools jumped from 66% in 2024 to 92% in 2025, while the share using GenAI specifically for assessments climbed from 53% to 88% over the same period (Freeman, 2025). Elsewhere the pattern is much the same. The Digital Education Council's 2024 Global AI Student Survey, drawing on 3,839 responses from bachelor's, master's, and doctoral students across sixteen countries, put overall AI adoption at 86%, with ChatGPT the clear leader at 66% usage. Chegg's 2025 global survey of 11,706 undergraduates across fifteen countries arrived at a similar figure, 80% of students worldwide having used GenAI to support their coursework. Institution-level data tell the same story: 87.5% of Harvard undergraduates by spring 2024, 85% of a broader U.S. sample by 2025, and 71.4% of higher-education students globally by early that same year. Different samples, different survey instruments, roughly the same number - which suggests this is a genuine, fast-moving phenomenon rather than an artifact of any one study's methodology.

None of this settles the more interesting question underneath it, though: using a tool is not the same thing as being literate in it, and being literate in it is not the same thing as benefiting cognitively from it. The very same 2024 Digital Education Council survey that found 86% adoption also found that 58% of students felt they lacked sufficient AI knowledge and skills, and 48% felt unprepared for an AI-shaped workforce. That gap between how often students use these tools and how competent they actually feel using them is really the starting point for this review.

Universities exist, in no small part, to build what Bloom's original taxonomy and its later revision call higher-order cognitive processes: analysis, evaluation, and creation, as opposed to the lower-order work of remembering, understanding, and routine application (Anderson & Krathwohl, 2001). Bissell and Lemons (2006) showed years before any of this that deliberately building assessment around these higher tiers improves not just content mastery but students' awareness of their own reasoning. The worry with GenAI is that a tool capable of producing analysis-shaped and creation-shaped output on a student's behalf could hollow out exactly the processes universities are supposed to be cultivating.

The evidence on this worry is genuinely split down the middle. On one side sits a substantial body of work documenting real cognitive risk. Gerlich (2025), surveying and interviewing 666 people across age groups, found a significant negative correlation between frequent AI tool use and critical thinking scores, with cognitive offloading, the habit of pushing memory and reasoning tasks onto a tool rather than doing them oneself, acting as the mediator. Younger, heavier users in that study scored measurably worse on critical thinking

than older, lighter ones. Zhai et al. (2024), reviewing the over-reliance literature systematically, reached much the same conclusion about habitual dependence on AI-generated content. Fan et al. (2025) coined the phrase "metacognitive laziness" for what happens when learners start offloading not just tasks but the monitoring of their own learning, a pattern that also turned up as the leading risk category in 33.7% of the 89 studies Alubthane (2026) reviewed.

On the other side sits an equally substantial body documenting real benefit, just under different conditions. Urban et al. (2024), running a proper randomized experiment, found that ChatGPT use improved solution quality, elaboration, and originality among university students working on creative problem-solving. Zhou et al. (2025) showed that ChatGPT-facilitated scaffolding measurably improved mathematical problem-solving, and Chen et al. (2025) found that argument-map-supported ChatGPT use sharpened critical reasoning in argumentative writing tasks. Meng et al. (2025) found that a GenAI-literacy-integrated instructional approach lifted both creative and critical thinking among pre-service teachers.

What looks like a contradiction between these two bodies of evidence turns out, on closer inspection, not to be much of one. The relationship between GenAI use and HOTS is conditional, not direct. Alubthane (2026) formalised this as a "dual-mechanism model": GenAI works as a cognitive amplifier under structured, metacognitively guided conditions, and as a cognitive substitute under unguided, passive ones. Tellingly, only 44.9% of the 89 studies in that review reported any specific pedagogical strategy for GenAI integration, which means that for most of the documented classroom use out there, whether GenAI amplifies or replaces student thinking was left largely to chance rather than to instructional design.

The variable that keeps surfacing as the explanation for this split is self-regulation. Zhou et al. (2024) found that self-regulated learning significantly mediates the relationship between GenAI use and both critical thinking and problem-solving among university students. Sardi et al. (2025), reviewing this relationship specifically, concluded that students with stronger self-regulated learning capacities collaborate with GenAI more productively, while those with weaker SRL lean toward unstructured, dialogue-driven reliance on the tool. Shi et al. (2025), surveying 257 Chinese university students, found that both AI literacy and SRL positively predicted writing performance, with SRL showing the stronger of the two effects - which hints that how a student regulates their own learning may matter more than how technically fluent they are with the tool itself.

Zimmerman's (2002) influential model frames self-regulated learning as a cyclical process running through three phases: forethought (goal-setting and strategic planning), performance (strategy execution and self-monitoring), and self-reflection (evaluating outcomes against the original goals), all grounded in social cognitive theory (Panadero, 2017). That cyclical structure gives a natural bridge to the GenAI literacy question. A student who sets a goal before querying a GenAI tool, actually monitors the output rather than accepting it wholesale, and reflects afterward on what was genuinely learned versus merely retrieved is doing something quite different cognitively from a student who treats the tool as a black box that spits out answers.

Even with all this converging evidence, three gaps remain. First, most existing work treats GenAI literacy as one-dimensional, usually operationalised as frequency or self-reported comfort of use, rather than engaging with the multidimensional frameworks the AI-literacy literature has itself proposed, which separate out technical, critical-evaluative, ethical, and metacognitive competencies (Kassorla et al., 2024; Zhou & Schofield, 2024). Second, while SRL's mediating role has been demonstrated empirically (Zhou et al., 2024; Shi et al., 2025), few reviews have systematically mapped specific AI literacy dimensions onto specific phases of Zimmerman's SRL cycle, so the mechanism remains somewhat black-boxed even where the outcome is well established. Third, the pedagogical and curricular implications of all this evidence are scattered across discipline-specific case studies, business education here, nursing

there, information systems elsewhere, rather than pulled together into general design principles a lecturer or curriculum committee could actually use.

This narrative review is organised around three research questions:

RQ1. What dimensions of generative AI literacy appear in the current literature, and how does each dimension relate, conceptually and empirically, to the development of Higher Order Thinking Skills among university students?

RQ2. How does Self-Regulated Learning function as the mechanism linking generative AI literacy to Higher Order Thinking Skills, and through which phases of the SRL cycle does this linkage actually operate?

RQ3. What do these findings imply for pedagogical design and curriculum development in higher education, particularly for structuring GenAI integration so that it supports rather than substitutes for higher-order cognitive development?

2. METHODS

2.1 Research Design

This article takes a narrative review approach rather than a systematic review or meta-analysis. That choice follows from the questions themselves: RQ1 through RQ3 require pulling together several distinct but related bodies of literature, AI literacy frameworks, cognitive taxonomy theory, self-regulated learning theory, and empirical education research, in a way a single systematic search protocol would not comfortably accommodate. A narrative review allows the thematic organisation needed to connect theory (Bloom's Taxonomy, Zimmerman's SRL model) with empirical findings and then translate both into curricular recommendations, following the methodological precedent set out by Popay et al. (2006) for narrative synthesis.

2.2 Literature Search Approach

Literature was identified through repeated searches of databases and search engines indexing Scopus- and Web of Science-listed journals, including MDPI, ScienceDirect, Frontiers, Springer Nature, Taylor & Francis, and IEEE, supplemented by arXiv preprints where recent, directly relevant empirical work had not yet made it into an indexed journal, and by major institutional survey reports (HEPI, Digital Education Council, Chegg) for the adoption statistics used in the Introduction. Search terms combined variations of "generative AI literacy," "higher order thinking skills," "self-regulated learning," "critical thinking," "cognitive offloading," and "Bloom's taxonomy," run in stages matching each thematic pillar of the review. Consistent with the exploratory, non-systematic character of narrative review, the search was not restricted to one Boolean string executed once; it was refined repeatedly as gaps in the emerging picture became clear.

2.3 Source Selection Criteria

Sources were chosen with three considerations in mind: publication in a peer-reviewed, indexed journal, with a preference for work from 2023 onward to capture the post-ChatGPT literature while still retaining a handful of earlier foundational sources (Flavell, 1979; Zimmerman, 2002; Sweller, 1988) where these remain seminal and actively cited in current empirical work; direct relevance, empirical or conceptual, to at least one research question; and verifiable authorship, venue, and DOI or stable URL, checked at the point each source was retrieved. Institutional survey reports were included specifically to ground the Introduction's statistics in recent, verifiable primary data rather than secondary estimates.

2.4 Synthesis Approach

Literature is organised thematically here rather than study by study or chronologically, following the convention Popay et al. (2006) describe for narrative synthesis. Four thematic

clusters structure the Results and Discussion: the dimensional architecture of GenAI literacy; how HOTS is conceptually anchored in Bloom's Revised Taxonomy and relates to GenAI use; Zimmerman's SRL cycle as the mechanism connecting AI literacy to HOTS; and the pedagogical and curricular implications, including the specific risks, cognitive offloading and academic integrity, that curriculum design has to reckon with. Within each cluster, studies that agree are grouped together, and studies that disagree or report conditional findings are placed in direct contrast, which suits the more exploratory, dialectical character of a narrative review as opposed to the protocol-driven aggregation typical of systematic review.

3. RESULTS AND DISCUSSIONS

3.1 The Dimensional Architecture of Generative AI Literacy

The first thing that stands out in this literature is that "AI literacy" resists being treated as a single skill. Several independently developed frameworks converge on broadly similar structures despite using different labels. The EDUCAUSE framework organises AI literacy in teaching and learning around four areas: technical understanding, evaluative skills, practical application, and ethical considerations (Kassorla et al., 2024). The Digital Education Council's framework proposes five dimensions and leans explicitly human-centred, putting critical thinking, judgment, and emotional intelligence alongside technical understanding of AI and data (Digital Education Council, 2025). Zhou and Schofield (2024) offer a bottom-up curricular framework built around knowing and understanding AI, using and applying AI, evaluating and creating with AI, and navigating AI ethics, explicitly designed to move students from basic awareness toward critical, generative engagement.

What unites these frameworks, despite the different labels, is a shared insistence that technical operational competence, knowing how to prompt a tool and read its output, is necessary but not sufficient. Each treats critical evaluation and ethical judgment as distinct, higher-tier competencies that do not simply fall out of technical fluency. That distinction matters for RQ1: it suggests that "AI literacy" as usually measured in survey research (how often someone uses the tool, how comfortable they say they feel) captures, at best, the technical layer of a much broader construct, which may explain why frequency of GenAI use alone shows such an inconsistent relationship with HOTS across the studies discussed in Section 1.2. A student can be technically fluent, capable of writing sophisticated prompts and navigating the interface with ease, while remaining critically and metacognitively naive about what the tool actually returns.

The distinction holds up under direct empirical scrutiny, too. Tolzin et al. (2024), looking at AI literacy's role in prompt engineering, found that certain literacy dimensions specifically improve prompt quality and model adaptation, and argued for deliberately building AI-related content into curricula rather than assuming literacy simply develops from repeated exposure. A scoping review of prompt engineering in higher education reached a similar conclusion, identifying prompt literacy, guided learning design, and GenAI usage risk as separate thematic areas each requiring its own curricular attention, rather than one undifferentiated "AI skills" module.

3.2 Higher Order Thinking Skills: Conceptual Anchoring

Bloom's original taxonomy, and the 2001 revision by Anderson and Krathwohl, arranges cognitive objectives into six hierarchical levels: remember, understand, apply, analyze, evaluate, and create. The lower three, often shortened to LOTS (Lower-Order Thinking Skills), cover recall, comprehension, and routine application. The upper three, HOTS, require learners to break material into components, judge it against explicit criteria, and build something new out of existing elements (Anderson & Krathwohl, 2001). Crowe et al. (2008) relabel these same

tiers as LOCS and HOCS, but the underlying hierarchy is the same one that has anchored educational research on HOTS for decades, whatever the acronym.

How this connects to GenAI use is fairly direct, and arguably structural. GenAI tools happen to be extraordinarily good at operations that resemble the lower tiers of the taxonomy: retrieving, summarising, restating. Whether they can genuinely perform the upper tiers on a student's behalf is much more contested. Ask a GenAI tool to "explain" a concept, and it is doing understand-level work for the student. Ask it to analyze a framework's strengths and weaknesses and then accept the output without checking it, and the analyze- and evaluate-level cognitive work has been substituted rather than supported. Bissell and Lemons (2006) showed, back before GenAI existed, that deliberately designing assessment around HOTS-tier objectives, rather than assuming students will move beyond content-level questions on their own, measurably improves both critical-thinking performance and students' awareness of the reasoning being asked of them. That finding matters even more now: if assessment does not explicitly demand analyze-, evaluate-, and create-tier engagement, GenAI's fluency at LOTS-tier operations creates a built-in incentive for students to stop doing their own cognitive work right where HOTS begins.

Alubthane's (2026) systematic review offers the most granular evidence on this to date. Across 89 studies, critical thinking was the most frequently targeted HOTS domain (76.4%), followed by problem-solving (62.9%) and creative thinking (36.0%). Positive effects showed up in 40.4% of studies, negative effects in 16.9%, and mixed or conditional effects in 23.6%, a distribution that itself supports the conditional rather than uniform relationship this review is arguing for.

3.3 Self-Regulated Learning as the Mechanism Linking AI Literacy to HOTS

3.3.1 Zimmerman's Cyclical Model

Zimmerman's (2002) cyclical model, grounded in social cognitive theory, describes learning as a three-phase cycle: forethought (task analysis, goal-setting, strategic planning, self-motivational beliefs), performance (self-control and self-observation during task execution), and self-reflection (judging outcomes against goals and the self-reactions that shape future engagement) (Panadero, 2017). Each phase depends on the kind of metacognitive monitoring and control Flavell (1979) first described as the active regulation of one's own thought processes.

This model gives a useful lens for understanding why the same GenAI tool produces such different outcomes across different students. In the forethought phase, a self-regulated student approaches a GenAI query with an actual learning goal in mind, using the tool as a scaffold for understanding something before analyzing it independently, rather than treating it as a shortcut to a finished product. In the performance phase, that same student monitors the output actively, checking claims, noticing inconsistencies, treating the response as a draft rather than a final answer. In the self-reflection phase, they evaluate what was actually learned versus what was simply retrieved, and adjust their approach next time.

3.3.2 Empirical Evidence for SRL's Mediating Role

This mapping holds up empirically. Zhou et al. (2024) found that self-regulation significantly mediates the relationship between GenAI use and both critical thinking and problem-solving among university students, meaning GenAI use predicts HOTS gains mostly to the extent it runs through self-regulatory processes, rather than producing gains on its own. Sardi et al. (2025), reviewing the SRL-GenAI-critical thinking literature systematically, reached a parallel conclusion: stronger self-regulation supports independent reasoning even in GenAI-assisted work, while weaker self-regulation shows up as the kind of unstructured, dialogue-driven engagement Zhou et al. (2024) elsewhere link to leaning on external help instead of thinking things through.

Shi et al. (2025), using structural equation modelling with 257 university students in China, found that both AI literacy and SRL positively predicted writing performance, with SRL the stronger predictor of the two, and also found AI literacy positively linked to GenAI-driven wellbeing, with writing performance partly mediating that link. It's a useful finding because it suggests the benefit of AI literacy may run partly through wellbeing and partly through direct skill application, but in either case self-regulation remains the stronger predictor, which reinforces the case that SRL, not raw AI literacy, is doing most of the actual work in determining HOTS-relevant outcomes.

Similar patterns show up outside writing specifically. In programming tasks, students with stronger self-regulated learning skills collaborated more effectively with GenAI, while those with weaker SRL leaned toward overdependence. A related distinction between "mastery-oriented" and "procedural" approaches to GenAI use found that students taking a mastery orientation, using AI to build and extend their own understanding, ended up with stronger academic performance and more developed critical thinking than students who simply reproduced AI-generated content unmodified.

3.3.3 The Amplifier-Substitute Divergence Explained Through SRL

Putting these pieces together, Alubthane's (2026) dual-mechanism model can be reframed usefully through Zimmerman's cycle: GenAI acts as a cognitive amplifier exactly when a student's engagement with it is scaffolded by forethought, monitored during performance, and reflected on afterward, and acts as a cognitive substitute exactly when those SRL phases are missing or weak. That reframing gives some precision to what might otherwise look like a simple dosage effect, more AI use is worse, or more AI use is better. The variable that actually matters is not how much a student uses GenAI, but which phase of the SRL cycle, if any, is governing that use.

Direct evidence for this comes from Hong et al. (2025), who built a quasi-experimental study around what they called "cognitive offload instruction", deliberately handing lower-order tasks to GenAI while explicitly requiring evaluative engagement from students on the higher-order components, and found this produced greater critical thinking gains than conventional instruction. That result matters because it shows offloading itself is not inherently harmful. The very mechanism that constitutes a documented risk under unguided conditions (Zhai et al., 2024; Fan et al., 2025) becomes a genuine pedagogical asset once it sits inside an SRL-consistent instructional design that preserves forethought and reflection around whatever tasks remain the student's own to do.

3.4 The Cognitive Risk Evidence: Cognitive Offloading, Over-Reliance, and Academic Integrity

A complete account of the AI literacy-HOTS relationship has to engage honestly with the risk literature; curricular recommendations that ignore documented harm would not be worth much. Gerlich's (2025) mixed-methods study of 666 participants found a significant negative correlation between AI usage frequency and critical thinking scores, mediated by cognitive offloading, with younger participants showing both higher dependency and lower critical thinking scores than older ones. Interestingly, higher educational attainment was independently associated with stronger critical thinking regardless of AI usage level, which suggests general academic preparation may buffer against offloading risk to some extent, though it clearly does not eliminate it.

A related but distinct strand of evidence examines what one line of research calls the "double-edged sword" quality of information literacy under GenAI conditions: literate students, precisely because they monitor and cross-reference AI output so carefully, may end up more cognitively fatigued from that verification burden, even as the same vigilance protects them from the passive acceptance that defines cognitive offloading in less literate users. This

complicates any simple claim that more literacy is costless, and points toward a real need for verification strategies that are efficient as well as rigorous, which is itself an argument for teaching explicit SRL strategies rather than assuming literacy alone will sort out the tension.

Academic integrity is the second major risk domain in this literature. A systematic review synthesising 41 studies on GenAI and academic integrity in higher education concluded that while GenAI can genuinely enhance engagement and efficiency, it also carries significant risk of academic dishonesty that institutional policy has been slow to catch up with. Survey evidence from U.S. university students found that students' own ethical beliefs, not their awareness of institutional policy, were the strongest predictors of both perceived misconduct and actual AI-assisted writing behaviour, with policy awareness showing no significant independent effect. That has a direct pedagogical implication: policy alone, without explicit ethical and metacognitive instruction alongside it, does not appear to be enough to shape how students actually behave around GenAI, which reinforces the broader argument here that the ethical dimension of AI literacy needs to be actively taught rather than communicated through a set of guidelines nobody reads closely.

3.5 Implications for Pedagogical Design and Curriculum Development

Taken together, the dimensional architecture of AI literacy, the HOTS framework, the SRL mechanism, and the documented risk profile point toward a few converging implications for curriculum design, which speak directly to RQ3.

AI literacy and SRL should be taught as companion competencies, named explicitly, rather than assumed to develop on their own from tool exposure. The consistent finding that self-reported usage frequency does not reliably predict HOTS outcomes, while SRL consistently does (Zhou et al., 2024; Shi et al., 2025), suggests that curricula focused only on technical tool training, how to prompt, how to navigate an interface, are aiming at the wrong variable. Scaffolded prompting instruction, systematic verification habits, and built-in reflective practice, as demonstrated in inquiry-led models of GenAI integration in information systems education, offer a workable template: courses that move students deliberately along an arc from dependency to collaboration to integration produce more durable gains than courses that simply allow tool access and hope for the best.

Assessment design needs to target the analyze, evaluate, and create tiers of Bloom's taxonomy explicitly, since GenAI's fluency with the lower tiers otherwise gives students every reason to stop short of HOTS-level engagement. This echoes Bissell and Lemons's (2006) pre-GenAI finding about HOTS-oriented assessment, and it is reinforced by the finding that structured pedagogical strategies, reflective learning, collaborative learning, scaffolding, argumentation-based feedback, were associated with more consistently positive HOTS outcomes than unstructured GenAI exposure across the 89-study corpus Alubthane (2026) reviewed.

Curricula should also deliberately build in the "cognitive offload instruction" principle Hong et al. (2025) demonstrated: assigning lower-order tasks, information retrieval, first drafts, routine summarising, to GenAI on purpose, while explicitly reserving evaluative, analytical, and creative work for the student. That reframes offloading from an incidental risk into a deliberate design choice, provided it comes with the SRL-consistent monitoring and reflection this review has argued is the real determinant of whether offloading amplifies or replaces student cognition.

The ethical and academic-integrity side of AI literacy needs active instruction rather than policy announcements, given that policy awareness alone does not predict behaviour while ethical belief does. Case-based ethical reasoning built into coursework, rather than a one-time policy email, looks like a more promising lever for shaping responsible use.

Institutions, finally, need to take seriously how genuinely multidimensional AI literacy is, and resist collapsing curriculum design into technical-skills training alone. The convergence

of the EDUCAUSE, Digital Education Council, and Zhou and Schofield (2024) frameworks around distinct technical, evaluative, ethical, and metacognitive dimensions suggests a single "how to use ChatGPT" workshop cannot substitute for staged, deliberate development across all four.

3.6 Limitations of the Existing Evidence Base

A few limitations in this literature bear directly on how much confidence to place in the implications above. The empirical corpus remains geographically lopsided: Alubthane's (2026) review found 67.4% of studies conducted in multi-national or unspecified contexts and only 19.1% representing the Global South, with very little from Southeast Asia or the Arab world, a gap that matters directly for non-Western higher education systems, Indonesian universities among them, where local empirical evidence on the GenAI-SRL-HOTS relationship remains thin. Longitudinal designs are almost entirely absent; most studies rely on cross-sectional, single-timepoint measurement, leaving open whether SRL-mediated GenAI benefits hold up, strengthen, or fade over the course of a full degree programme. Explicit theoretical grounding is also uncommon: only 25.8% of the 89 studies Alubthane (2026) reviewed used an identifiable theoretical framework, which limits how cumulative and comparable knowledge-building in this field can really be.

4. CONCLUSION

This narrative review set out to examine how the dimensions of generative AI literacy relate to university students' Higher Order Thinking Skills, and how Self-Regulated Learning governs that relationship. Three conclusions answer the questions posed at the outset.

On RQ1, the literature treats GenAI literacy as multidimensional rather than unitary, spanning technical, critical-evaluative, ethical, and metacognitive competencies (Kassorla et al., 2024; Digital Education Council, 2025; Zhou & Schofield, 2024). That structure connects directly to HOTS because GenAI tools disproportionately automate the lower tiers of Bloom's taxonomy, which means technical literacy alone reliably supports only lower-order engagement, while it is the critical-evaluative and metacognitive dimensions that actually determine whether analyze-, evaluate-, and create-tier engagement survives.

On RQ2, the evidence points to Self-Regulated Learning, worked through Zimmerman's (2002) three-phase cycle, as the proximate mechanism connecting AI literacy to HOTS outcomes, with direct empirical mediation shown by Zhou et al. (2024) and backed up by Shi et al. (2025) and Sardi et al. (2025). The same GenAI engagement amplifies thinking when it is governed by forethought, active monitoring, and reflection, and substitutes for it, through offloading, metacognitive laziness, over-reliance, when those SRL phases are weak or missing (Gerlich, 2025; Fan et al., 2025; Zhai et al., 2024).

On RQ3, these findings suggest higher education institutions should treat AI literacy and SRL as competencies to be taught together and explicitly, embedded in curriculum and assessment design, HOTS-targeted assessment, deliberate cognitive offload instruction, active ethical case-based teaching, rather than as incidental byproducts of simply giving students access to GenAI tools.

The theoretical contribution here lies mainly in reframing the GenAI-cognition debate away from a binary helps-or-harms question and toward a conditional model where SRL moderates and mediates the direction of effect, a reframing that extends Alubthane's (2026) dual-mechanism model by specifying which SRL-cycle phases actually carry amplification versus substitution.

Practically, this points toward specific design moves: teaching AI literacy alongside explicit SRL strategy instruction, redesigning assessment toward HOTS-tier demands GenAI cannot easily substitute for, deliberately deciding which cognitive tasks get offloaded to GenAI and which stay with the student, and addressing academic integrity through active ethical instruction rather than policy statements alone.

Some limitations of this review itself are worth naming. As a narrative rather than systematic review, source selection followed consistent criteria but not a fully reproducible PRISMA-style protocol, and the synthesis reflects thematic organisation rather than exhaustive corpus coverage. The underlying evidence base, as discussed in Section 3.6, remains geographically skewed and mostly cross-sectional.

Future research might usefully pursue longitudinal studies tracking the SRL-GenAI-HOTS relationship across a full degree programme rather than a single course; empirical work situated specifically in Southeast Asian and Indonesian higher education, where institutional, linguistic, and cultural conditions may shift the relationships documented here in predominantly Western, Chinese, and Middle Eastern samples; intervention studies testing the cognitive offload instruction model Hong et al. (2025) proposed across disciplines beyond English writing; and integrated instruments that measure AI literacy's dimensions and SRL phases jointly, rather than as separately measured constructs, extending the kind of structural modelling Shi et al. (2025) began but which remains rare in this literature.

4. REFERENCES

- Abdallah, N., Katmah, R., Khalaf, K., & Jelinek, H. F. (2025). Systematic review of ChatGPT in higher education: Navigating impact on learning, wellbeing, and collaboration. *Social Sciences & Humanities Open*, 12, 101866. <https://doi.org/10.1016/j.ssaho.2025.101866>
- Agaoglu, F. O., Bas, M., Tarsuslu, S., Ekin, L. O., & Agaoglu, N. B. (2025). The mediating digital literacy and the moderating role of academic support in the relationship between artificial intelligence usage and creative thinking in nursing students. *BMC Nursing*, 24, 484. <https://doi.org/10.1186/s12912-025-03128-3>
- Alubthane, F. O. (2026). Amplifier or substitute? A systematic review of generative AI's impact on higher-order cognitive skills among university students. *Frontiers in Psychology*, 17, 1863931. <https://doi.org/10.3389/fpsyg.2026.1863931>
- Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives (Complete ed.). Addison Wesley Longman.
- Chen, X., Jia, B., Peng, X., Zhao, H., Yao, J., Wang, Z., et al. (2025). Effects of ChatGPT and argument map (AM)-supported online argumentation on college students' critical thinking skills and perceptions. *Education and Information Technologies*, 30, 17623-17658. <https://doi.org/10.1007/s10639-025-13471-2>
- 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, 865-882. <https://doi.org/10.1080/03075079.2024.2316881>
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., et al. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56, 489-530. <https://doi.org/10.1111/bjet.13544>
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist*, 34(10), 906-911.
- Freeman, J. (2025). Student generative AI survey 2025. Higher Education Policy Institute. <https://www.hepi.ac.uk/reports/student-generative-ai-survey-2025/>
- Gammoh, L. A. (2024). ChatGPT in academia: Exploring university students' risks, misuses, and challenges in Jordan. *Journal of Further and Higher Education*, 48, 608-624. <https://doi.org/10.1080/0309877X.2024.2378298>
- 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>

- Hong, H., Vate-U-Lan, P., & Viriyavejakul, C. (2025). Cognitive offload instruction with generative AI: A quasi-experimental study on critical thinking gains in English writing. *Forum for Linguistic Studies*, 7, 325-334. <https://doi.org/10.30564/fls.v7i7.10072>
- Hou, C., Zhu, G., Sudarshan, V., Lim, F. S., & Ong, Y. S. (2025). Measuring undergraduate students' reliance on generative AI during problem-solving: Scale development and validation. *Computers & Education*, 234, 105329. <https://doi.org/10.1016/j.compedu.2025.105329>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kassorla, M., Georgieva, M., & Papini, A. (2024, October 17). AI literacy in teaching and learning. *EDUCAUSE*. <https://www.educause.edu/content/2024/ai-literacy-in-teaching-and-learning/introduction>
- Meng, X.-P., Guo, X. G., Fang, J. W., Chen, J., & Huang, L.-F. (2025). Fostering pre-service teachers' generative AI literacy and critical thinking: An RSCQA approach. *Educational Technology & Society*, 28, 202-225. [https://doi.org/10.30191/ETS.202507_28\(3\).TP01](https://doi.org/10.30191/ETS.202507_28(3).TP01)
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, 422. <https://doi.org/10.3389/fpsyg.2017.00422>
- Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., et al. (2006). Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC methods programme. Lancaster University. <https://doi.org/10.13140/2.1.1018.4643>
- Sardi, J., Candra, O., Yuliana, D. F., Yanto, D. T. P., & Eliza, F. (2025). How generative AI influences students' self-regulated learning and critical thinking skills? A systematic review. *International Journal of Engineering Pedagogy*, 15, 94-108. <https://doi.org/10.3991/ijep.v15i1.53379>
- Shi, J., Liu, W., & Hu, K. (2025). Exploring how AI literacy and self-regulated learning relate to student writing performance and well-being in generative AI-supported higher education. *Behavioral Sciences*, 15(5), 705. <https://doi.org/10.3390/bs15050705>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4
- Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., et al. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education*, 215, 105031. <https://doi.org/10.1016/j.compedu.2024.105031>
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11, 28. <https://doi.org/10.1186/s40561-024-00316-7>
- Zhou, X., Teng, D., & Al-Samarraie, H. (2024). The mediating role of generative AI self-regulation on students' critical thinking and problem-solving. *Education Sciences*, 14(12), 1302. <https://doi.org/10.3390/educsci14121302>
- Zhou, H., Chen, Y., Liu, Y., Jiang, R., Wang, J., & Sun, M. (2025). Harnessing generative AI and argumentation-driven learning for entrepreneurial competence development: Evidence from university-based studies. *Interactive Learning Environments*, 34, 1136-1153. <https://doi.org/10.1080/10494820.2025.2519122>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64-70. https://doi.org/10.1207/s15430421tip4102_2