Indonesian Journal of Innovation and Applied Sciences (IJIAS)
Journal Homepage: https://ojs.literacyinstitute.org/index.php/ijias
E-ISSN: 2775-4162

Volume 6	Issue 2	June (2026)	DOI: 10.47540/ijias.v6i2.2774	Page: 271 – 282

The GPS for thinking: How AI partners are reshaping student metacognition

Sayed Mahbub Hasan Amiri[1], Naznin Akter[2], Marzana Mithila[3], Md. Mainul Islam[1]
[1]Department of CS, Dhaka Residential Model College, Bangladesh
[2]Department of English, Shamlapur Ideal Academy, Bangladesh
[3]Department of Field Work, Unique Personnel (UK) Limited, United Kingdom
Corresponding Author: Sayed Mahbub Hasan Amiri; Email: amiri@drmc.edu.bd

ARTICLE INFO

Keywords: Artificial Intelligence, Cognitive Offloading, GPS Metaphor, Metacognition, Self-Regulated Learning.

Received : 02 May 2026
Revised : 05 June 2026
Accepted : 30 June 2026

ABSTRACT

Generative artificial intelligence is rapidly becoming a cognitive partner in education, capable of planning tasks, monitoring progress, and evaluating solutions on a learner's behalf. This conceptual synthesis paper examines the risk that such AI tools, while improving immediate performance, may erode students' metacognitive abilities in their capacity to plan, monitor, and evaluate their own thinking. Drawing a parallel with GPS navigation research, where habitual turn-by-turn guidance has been shown to impair spatial memory and hippocampal engagement, we introduce the metaphor of AI as a "GPS for thinking". Through an integrative review of literature spanning cognitive psychology, neuroscience, and the learning sciences, we synthesize evidence that AI-assisted learning can lead to a form of cognitive disuse atrophy, specifically by short-circuiting the metacognitive loop. Emerging studies reveal that students who rely heavily on AI tutors often perform worse when the tool is removed, suffer from an illusion of explanatory depth, and struggle to articulate the reasoning behind their answers. To counter these effects, we propose a shift from a GPS model where the tool issues commands to a compass model, where the tool provides orientation while preserving learner agency. Five evidence-informed design principles are advanced: prompting planning before assistance, delaying and fading feedback, embedding mandatory reflection pauses, making AI reasoning visible, and calibrating learners' confidence. The article argues that the long-term goal of educational AI must be to strengthen, not supplant, the student's inner compass.

INTRODUCTION

Consider a sixteen-year-old student, Lena, working on a physics problem about conservation of momentum in a bright classroom in early 2026. She types the question into an AI tutoring interface. Within seconds, the screen fills with a perfectly ordered sequence of steps: identify the system, define initial and final states, apply the conservation equation, solve for the unknown velocity. Lena watches, nods along, and transfers the final answer to her worksheet. When her teacher kneels beside her and asks, "Can you explain how you solved it?", Lena falters. She can repeat the steps mechanically, but she cannot explain why she began with that particular equation, what alternative paths she might have taken, or how she would handle a differently worded problem. She has arrived at the correct answer, but she holds no mental map of the intellectual terrain she just crossed. As noted in the OECD's (2021) comprehensive review, the rapid integration of AI into classrooms brings both opportunities and risks for deeper learning processes.

This scenario, replicated in millions of classrooms around the world, captures a fundamental tension of the artificial intelligence era in education. Generative AI large language models and their multimodal successors have moved from futuristic curiosity to everyday classroom infrastructure with astonishing speed. Intelligent tutoring systems, writing assistants, and problem-solving companions now promise to make expertise widely available, personalize learning, and reduce persistent achievement gaps. Their capacity to break down complex tasks, offer immediate feedback, and generate clear explanations can lighten the cognitive load for novice learners, opening up rigorous content to students who might otherwise be left behind. However, as these tools become more powerful and more deeply integrated, a quiet but urgent question grows louder: when we hand over the moment-by-moment navigation of a learning task to an external intelligence, what happens to the learner's own ability to navigate?

This question is deeply rooted in half a century of cognitive science. Researchers have long identified metacognition, the capacity to plan, monitor, and evaluate one's own thinking as one of the most powerful predictors of academic achievement. First defined by Flavell (1979) as "cognition about cognitive phenomena," metacognition acts as the mind's internal guidance system, the quiet inner voice that asks: What do I already know? Is this making sense? Should I try a different strategy? Did that work? When students think metacognitively, they construct what can be called a cognitive map of the problem space: a mental representation that helps them orient themselves not only in the current task but also in unfamiliar future ones. Schraw & Moshman (1995) further distinguish among tacit, informal, and formal metacognitive theories, highlighting how learners' beliefs about their own cognition shape strategy use. Hattie's (2009) landmark synthesis of over 800 meta-analyses ranked metacognitive strategies among the most impactful educational interventions known, with an effect size comparable to that of prior academic ability itself. However, metacognition is not automatic. It develops slowly, through deliberate practice, in environments that require learners to articulate their reasoning, grapple with confusion, and make strategic choices (Azevedo & Gaševic, 2019). Recent meta-analyses confirm that metacognitive scaffolding within intelligent tutoring systems produces meaningful gains in student learning outcomes (Chen & Chavez, 2022). It is, in other words, precisely the kind of cognitive function most vulnerable to being quietly replaced by an obliging AI. Veenman and colleagues (2006) emphasize that metacognitive skills are domain general but require domain-specific knowledge to be effectively applied, a nuance often overlooked in AI learning designs.

The danger becomes strikingly clear when viewed through an unexpected but revealing parallel: the way satellite navigation technology reshapes the human brain. Over the past twenty years, cognitive neuroscientists have built a strong body of evidence showing that habitual reliance on GPS devices degrades the neural and cognitive systems that support spatial navigation (Dahmani & Bohbot, 2017; Ishikawa et al., 2008). In a key study, participants who navigated an unfamiliar environment using a GPS-based system reached their destinations efficiently but later produced much poorer sketch maps than those who used paper maps or direct experience (Ishikawa et al., 2008). They had moved through space without building a mental model of it. Later neuroimaging research confirmed that frequent GPS users have reduced grey matter volume in the hippocampus and perform worse on self-guided navigation tasks (Dahmani & Bohbot, 2017). The mechanism at work is a form of cognitive disuse atrophy: when an external tool consistently performs a cognitive function, the brain, ever efficient, reduces its investment in that function (Risko & Gilbert, 2016). London taxi drivers, who famously enlarge their posterior hippocampi through years of mapless navigation (Maguire et al., 2006), show the opposite principle: what you use, you build; what you outsource, you lose. The principles of cognitive load theory and deliberate practice, as synthesized in Kirschner & Hendrick (2020), underscore why offloading metacognitive processes to AI can be counterproductive.

Generative AI, we argue, presents education with a cognitive offloading challenge of a similar kind and even greater scope. Where GPS offloads spatial reasoning, AI offloads cognitive reasoning, the very processes of planning, monitoring, and evaluating that make up metacognition. When a student like Lena receives a solution that has already been broken into logical steps, her own planning is made unnecessary. When an AI silently corrects her mistakes in real time, her own monitoring is bypassed. When the final answer appears polished and accompanied by a smooth explanation, the reflective evaluation that solidifies learning is easily skipped. The AI becomes a GPS for thinking, navigating the intellectual landscape on the student's behalf. Moreover, the risk, as the navigation literature would predict, is not simply that students will become dependent on the tool in the moment, but that their internal cognitive mapping capacity and their metacognitive skills will weaken over time. The evolution from rule-based intelligent tutoring systems to generative AI companions has dramatically expanded both the capabilities and the potential cognitive costs of educational technology (Roll & Wylie, 2016).

Early research on AI in education already provides empirical support for this concern. A recent study found that high school students who practiced mathematics with a ChatGPT-based tutor outperformed their peers. At the same time, the AI was available but scored significantly lower on a later unassisted test, leading the researchers to conclude that generative AI without safeguards can harm learning (Bastani et al., 2025). Studies in computing education show a similar pattern: students who use AI code generators can describe what the generated code does but struggle to explain the strategic reasoning behind their approach, a clear sign of breakdown in metacognitive monitoring and evaluation (Prather et al., 2024). These results align with cognitive research showing that learners often mistake the fluency of an external explanation for their own understanding, an "illusion of explanatory depth" (Rozenblit & Keil, 2002) that AI's smooth, confident prose may uniquely strengthen. Roelle et al. (2017) showed that such fluency experiences can inflate confidence while leaving actual comprehension unchanged, directly undermining the self-monitoring component of metacognition. The emerging picture is one of fragile performance improvements hiding deeper developmental costs.

However, the appropriate response to this evidence is not to abandon AI. The GPS analogy, properly understood, does not condemn navigation technology; it condemns unexamined navigation technology that asks nothing of the user. Some cities have redesigned navigation interfaces to show compass orientation, provide whole route previews, and prompt drivers to identify landmarks, small acts of cognitive friction that keep the hippocampus engaged. The same principle can and should guide the design of AI learning partners. The aim is to build tools that support metacognition rather than replace it, turning the AI from a GPS that gives orders into a compass that offers direction while leaving the learner in charge of the journey.

The purpose of this research is threefold. First, to analyze how AI offloading affects student metacognition by reviewing the cognitive structure of metacognition and the ways current AI tools bypass its core elements. Second, to learn from the cognitive effects of GPS navigation, especially the framework of cognitive disuse atrophy, and apply that framework to intellectual learning. Third, to develop evidence-based design principles for AI that strengthen rather than erode metacognitive skills, and to present a "compass model" of educational AI. Accordingly, the paper is structured as follows. Section 2 reviews the interdisciplinary literature across four domains: GPS and spatial cognition, metacognition and academic achievement, AI in education and cognitive disuse, and design principles for metacognitive scaffolding. Section 3 details the integrative review methodology. Section 4 presents the synthetic framework, including the compass model and five design principles, and discusses implications for practice, policy, and future research. Section 5 addresses limitations and offers recommendations. A concluding section restates the central argument: that in learning, as in navigation, the destination matters less than the map, and that the most important thing an educational tool can do is leave the learner with a stronger internal compass.

METHODS

The methodology is presented below in four main sections as requested by the editor: Research Design, Informant Selection, Data Collection, and Data Analysis. Each section is further broken down into numbered points for clarity and transparency.

Research Design

This study is designed as an integrative literature review (Torraco, 2005; Snyder, 2019). This method is particularly suited to examining emerging, interdisciplinary topics where empirical findings from disparate fields must be synthesized into a novel theoretical framework.

The intersection of generative AI, cognitive offloading, and metacognition is nascent, fragmented, and conceptually diverse, drawing on cognitive neuroscience, educational psychology, human-computer interaction, and learning design. An integrative review is therefore more appropriate than a traditional systematic review, which is optimized for well-defined, mature domains with large bodies of homogeneous empirical studies (Torraco, 2005). Recent reviews of digital learning environments confirm that metacognitive prompts are effective only when they require active response rather than passive viewing (Schrader, 2020).

Unlike systematic reviews that aim for effect-size estimation, integrative reviews permit the inclusion of both empirical and theoretical literature (Whittemore & Knafl, 2005). This enables the paper to juxtapose well-established findings from spatial cognition (e.g., the neurological consequences of GPS use) with emerging but still preliminary evidence on AI-assisted learning. The overarching goal of the review is theory building, specifically: the construction of a conceptual framework (the "compass model") and a set of design principles grounded in cross-domain pattern recognition. The research design incorporates three integrated analytical strategies: thematic synthesis, comparative cross-domain analysis, and metaphor-driven theory development, as detailed in section 3.4.

Informant Selection

Informants (sources) were selected based on a systematic search and screening process guided by explicit inclusion and exclusion criteria.

Inclusion criteria were: (a) peer-reviewed empirical studies, meta-analyses, or major theoretical reviews published in English; (b) publication dates between 2000 and 2026, with the exception of foundational theoretical works from earlier periods (e.g., Flavell, 1979); (c) relevance to at least one of the four thematic clusters (GPS/cognitive offloading, metacognition, AI in education, or design principles); and (d) for empirical studies, clear reporting of methods and findings.

Exclusion criteria comprised: opinion pieces without empirical or theoretical grounding; studies focused exclusively on non-cognitive outcomes such as motivation or affect without a metacognitive or cognitive component; and papers addressing AI in non-educational contexts (e.g., industrial automation) without transferable cognitive implications.

Grey literature, such as the Bastani et al. (2025) working paper, was included when widely cited and methodologically transparent, given the rapid pace of AI research outstripping traditional publication cycles.

The initial search yielded 647 records after duplicate removal. Title and abstract screening eliminated 392 records that were clearly off-topic or failed inclusion criteria. Full-text review of the remaining 255 records resulted in a final corpus of 78 informants (sources) that directly informed the synthesis. The final corpus comprised studies from spatial cognition (n = 14), metacognition and self-regulated learning (n = 23), AI in education (n = 18), and design principles and cognitive theory (n = 23). A PRISMA-style flow diagram (Moher et al., 2009) was maintained to document the selection process.

Data Collection

Data collection (literature searching) was conducted between January and March 2026 across five electronic databases selected for their disciplinary coverage: PsycINFO (cognitive and educational psychology), ERIC (education research), PubMed (neuroscience and cognitive science), Scopus (multidisciplinary), and Google Scholar (supplementary and grey literature).

Search terms were organized into four thematic clusters corresponding to the paper's core constructs:

1. Cluster 1 (GPS navigation and cognitive offloading): ("GPS" OR "satellite navigation" OR "turn-by-turn navigation") AND ("spatial memory" OR "hippocampus" OR "cognitive map" OR "cognitive offloading").
2. Cluster 2 (Metacognition): ("metacognition" OR "metacognitive" OR "self-regulated learning" OR "SRL") AND ("planning" OR "monitoring" OR "evaluation" OR "self-assessment").
3. Cluster 3 (AI in education): ("generative AI" OR "ChatGPT" OR "AI tutor" OR "intelligent tutoring system") AND ("learning outcomes" OR "cognitive offloading" OR "metacognition" OR "self-explanation").
4. Cluster 4 (Design principles): ("metacognitive scaffolding" OR "feedback timing" OR "confidence calibration" OR "reflection prompts") AND ("educational technology" OR "AI").

Boolean operators were used to combine clusters for the full search corpus. For example, the central integrative search combined cluster one terms with clusters two and three: ("GPS" OR "cognitive offloading") AND ("metacognition" OR "self-regulated learning") AND ("AI" OR "intelligent tutoring system").

Snowball sampling from reference lists of key papers supplemented the database searches, as did forward citation tracking of seminal works such as Ishikawa et al. (2008), Maguire et al. (2006), and Flavell (1979). For each included source, a standardized data extraction form was used to collect the following information: author(s) and year, disciplinary domain, study design (empirical, theoretical, review), key constructs investigated, principal findings, and relevance to the paper's guiding questions. The extraction form was piloted on a subset of ten papers and refined before full application.

Data Analysis

Data analysis proceeded through three interconnected phases: thematic synthesis, comparative cross-domain analysis, and metaphor-driven theory development.

Phase 1: Thematic synthesis: Following the procedures outlined by Thomas and Harden (2008), the analysis began with inductive coding of findings within each thematic cluster. For example, within the GPS navigation cluster, recurring descriptive themes included "hippocampal grey matter reduction", "sketch map impoverishment," "dose-response relationship with offloading", and "interface design as moderator". Within the AI-in-education cluster, themes included "performance decline on unassisted tasks," "dissociation between recognition and generation," "fluency-induced overconfidence", and "reflection prompts as a protective factor". These descriptive themes were then aggregated into analytical themes that cut across clusters, such as "cognitive disuse atrophy", "metacognitive bypass", and "strategic friction".

Phase 2: Comparative cross-domain analysis: The thematic outputs from the GPS and AI clusters were systematically compared to identify structural parallels and disanalogies. This comparative process was guided by the principle of analogical transfer (Gentner & Markman, 1997): the mapping of relational structures from a well-understood source domain (spatial navigation and the hippocampus) to a less-understood target domain (AI-assisted cognitive navigation and metacognition). Attention was paid to identifying the functional equivalents of key constructs: spatial planning mapped to cognitive planning, GPS turn-by-turn commands to AI step-by-step solutions, and hippocampal atrophy to metacognitive skill decline, while noting boundary conditions where the analogy does not hold. Meta-analytic evidence implicates prefrontal and parietal networks in metacognitive monitoring, suggesting that AI offloading may affect distributed executive control systems (Pozuelos & Moreno, 2021).

Phase 3: Metaphor-driven theory development: The final analytical phase employed the GPS-compass metaphor as an integrative device. Metaphors in scientific reasoning serve not merely as rhetorical ornament but as cognitive tools that structure problem representation and hypothesis generation (Lakoff & Johnson, 1980). The GPS metaphor was used diagnostically: it framed the problem by highlighting the structural similarity between offloading spatial navigation and offloading cognitive navigation. The compass metaphor was then developed prescriptively as a target image for redesigned AI tools that provide orientation without removing agency.

Derivation of design principles: The five design principles prompting planning before assistance, delaying and fading feedback, embedding mandatory reflection pauses, making AI reasoning visible, and calibrating learner confidence were not selected a priori. They emerged from the iterative interplay of the three analytical phases. For each principle, a specific metacognitive vulnerability identified in the AI literature (e.g., bypassed planning) was matched with a protective intervention from the self-regulated learning or feedback literature (e.g., planning prompts) and translated into a concrete AI design specification. The principles were then cross-checked against the GPS-compass framework to ensure they moved the AI's functionality from commanding to orienting.

Quality assurance in analysis: To ensure rigor, credibility was pursued through triangulation across disciplines. Transferability was addressed by providing a thick description of the theoretical constructs and design principles. Dependability and confirmability were supported by maintaining an audit trail of search strategies, inclusion decisions, and analytical memos, and by explicitly articulating the metaphorical lens that shapes the interpretation (Lincoln & Guba, 1985).

Quality appraisal: Quality appraisal of included sources was conducted using an adapted version of the Mixed Methods Appraisal Tool (MMAT; Hong et al., 2018) for empirical studies, supplemented by the AACODS checklist (Tyndall, 2010) for grey literature. Theoretical and review papers were assessed for conceptual clarity, internal consistency, and breadth of synthesis. No source was excluded solely on quality grounds; instead, quality ratings were used to weight the interpretive confidence assigned to different bodies of evidence. The concept of desirable difficulties, elaborated by Schmidt & Bjork (2017), explains why delaying feedback can strengthen long-term retention and metacognitive monitoring.

RESULTS AND DISCUSSION

Findings from GPS Navigation and Cognitive Offloading

1. F1: Habitual GPS use leads to efficient wayfinding during tool use but significantly impoverished sketch maps afterward (Ishikawa et al., 2008).
2. F2: Frequent GPS users show reduced hippocampal grey matter volume and lower hippocampal activity during self-guided navigation (Dahmani & Bohbot, 2017).
3. F3: A dose-response relationship exists: more frequent GPS use predicts poorer spatial memory performance (Dahmani & Bohbot, 2017).
4. F4: London taxi drivers who navigate without GPS exhibit enlarged posterior hippocampi (Maguire et al., 2006).
5. F5: The underlying mechanism is cognitive disuse atrophy: the brain reduces investment in functions consistently performed by external tools (Risko & Gilbert, 2016).
6. F6: Navigation interfaces with orientation cues, route previews, and decision points can mitigate hippocampal disengagement (Ishikawa et al., 2008; Gramann et al., 2017)

Findings from Metacognition Research

1. F7: Metacognition comprises metacognitive knowledge and metacognitive regulation (planning, monitoring, evaluation) (Flavell, 1979; Schraw & Dennison, 1994).
2. F8: Metacognitive strategies have a large effect size on academic achievement (Hattie, 2009).
3. F9: Metacognitive prompts in digital environments significantly improve learning outcomes (Zheng et al., 2019).
4. F10: Metacognition does not develop automatically; it requires deliberate practice in environments that prompt articulation and reflection (Azevedo & Gašević, 2019).
5. F11: Learners often overestimate their understanding of the "illusion of explanatory depth" (Rozenblit & Keil, 2002).
6. F12: Fluency of external explanations can inflate confidence without improving comprehension (Roelle et al., 2017).
7. F13: Self-explanation enhances learning by forcing metacognitive monitoring (Chi, 2000).

Findings from AI in Education and Cognitive Offloading

1. F14: Students using a ChatGPT-based tutor outperformed peers during AI-assisted practice but scored significantly lower on an unassisted assessment (Bastani et al., 2025).
2. F15: The negative effect was concentrated among the heaviest AI users, indicating a dose-response relationship (Bastani et al., 2025).
3. F16: Computing students who use AI code generators can describe generated code but struggle to articulate their own problem-solving strategies or debug independently (Prather et al., 2024).
4. F17: This dissociation between recognition and generation indicates bypassed metacognitive monitoring and evaluation (Prather et al., 2024).
5. F18: Generative AI without safeguards can harm independent learning, creating fragile performance gains that mask deeper costs (Bastani et al., 2025; Mollick & Mollick, 2023).
6. F19: The fluency of AI-generated explanations may uniquely amplify the illusion of explanatory depth (synthesis from Rozenblit & Keil, 2002; Roelle et al., 2017).

Findings from Design Principles for Metacognitive Scaffolding

1. F20: Prompting learners to articulate a plan before receiving assistance supports the planning phase of metacognition (Azevedo & Gaševic, 2019; Chi, 2000).

2. F21: Delayed feedback produces better long-term retention than immediate feedback (Butler & Winne, 1995).
3. F22: Fading support (worked examples ⟶ completion problems ⟶ independent practice) preserves desirable difficulty (Renkl & Atkinson, 2003; Bjork & Bjork, 2011).
4. F23: Structured reflection prompts increase metacognitive awareness and learning quality (Wise & Hsiao, 2019).
5. F24: Asking learners to rate confidence before receiving feedback improves calibration accuracy (Dunlosky & Rawson, 2012; Hacker et al., 2008).
6. F25: Making expert reasoning visible (e.g., decision trees) invites comparison and deeper monitoring (VanLehn, 2011; Aleven & Koedinger, 2002).

Phase 1: Thematic Synthesis

Aggregating the 25 findings points, four higher-order analytical themes emerge.

Analytical Theme 1: Cognitive Disuse Atrophy (F1, F2, F3, F5, F14, F15, F16). Consistent external offloading of a cognitive function leads to a measurable decline in internal capacity. In GPS users, this appears as hippocampal grey matter reduction and impoverished sketch maps. In AI users, it appears as impaired unassisted performance, inability to articulate reasoning, and a dose-response relationship.

Analytical Theme 2: Metacognitive Bypass (F7, F10, F16, F17, F19). Current AI tools often skip or replace the three core metacognitive phases: planning is pre-empted by step-by-step solutions; monitoring is rendered unnecessary by silent error correction; polished final explanations bypass evaluation. The dissociation between recognition and generation (Prather et al., 2024) directly operationalizes bypassed monitoring and evaluation.

Analytical Theme 3: Illusion of Competence (F11, F12, F18, F19). The fluency of AI-generated explanations creates a feeling of knowing that masks genuine comprehension gaps (Rozenblit & Keil, 2002). Roelle et al. (2017) showed that fluency inflates confidence without improving actual comprehension. Bastani et al. (2025) found that students who appeared competent during AI assistance later failed unassisted tests, a classic illusion of competence. The illusion of explanatory depth is particularly pronounced in digital learning environments where learners passively consume explanations (Wiley & Jensen, 2020), a risk amplified by fluent AI-generated prose.

Analytical Theme 4: Strategic Friction as Protective Factor (F6, F20F25). Design features that introduce "cognitive friction" planning prompts, delayed feedback, reflection pauses, visible reasoning, and confidence calibration preserve metacognitive engagement. Just as GPS interfaces with orientation cues mitigate hippocampal disengagement, educational AI tools that demand active learner input protect against metacognitive atrophy.

Phase 2: Comparative Cross-Domain Analysis

Using analogical transfer (Gentner & Markman, 1997), structural parallels between GPS and AI domains are mapped:

Table 1. Structural parallels between GPS and AI domains

Structural Element	GPS Domain (Source)	AI Domain (Target)
External tool function	Turn-by-turn navigation commands	Step-by-step solution steps
Cognitive function offloaded	Spatial navigation (path planning, landmark tracking)	Metacognitive reasoning (planning, monitoring, evaluation)
Neural substrate affected	Hippocampus (spatial memory)	Prefrontal networks (executive control) are hypothesized
Behavioral consequence during tool use	Efficient arrival, low effort	Correct answers, fluent explanation
Behavioral consequence after tool removal	Poor sketch maps, cannot retrace route	Poor unassisted performance, cannot explain reasoning
Long-term change	Hippocampal grey matter reduction	Hypothesized decline in metacognitive accuracy
Protective design feature	Orientation cues, route previews, decision points	Planning prompts, delayed feedback, reflection, visible reasoning, calibration

Key parallel: In both domains, performance during tool use is maintained or enhanced, creating an illusion of competence. Degradation becomes visible only when the tool is removed. This implies that standards in tool performance metrics are insufficient for evaluating educational AI; unassessed transfer and long-term retention measures are essential.

Boundary conditions: Spatial navigation is hippocampus-dependent, whereas metacognition involves distributed prefrontal and parietal networks. The analogy is functional (what is offloaded) rather than neuroanatomical.

Phase 3: Metaphor-Driven Theory Development, The Compass Model

Diagnostic use of the GPS metaphor. When an AI functions like a GPS, issuing commands, requiring no active planning, correcting errors silently, it turns the learner into a passive recipient, moving through the problem space without building a mental model.

Prescriptive use of the compass metaphor. A compass provides orientation ("north is that way") but requires the traveler to read the terrain, choose a route, and correct course independently. Transposed to AI learning partners, a compass-model AI provides hints rather than full solutions, demands learner plans, withholds immediate correction, inserts reflection pauses, reveals its own reasoning, and asks for confidence judgments.

Derivation of five design principles. From the intersection of analytical themes, the comparative mapping, and the compass metaphor, five evidence-informed design principles are derived (see Table 1). Interpretation of Table 2: The table synthesizes the five principles. Each principle targets a specific phase of the metacognitive loop (planning, monitoring, or evaluation) and addresses a primary risk (e.g., bypassed strategy selection, suppression of error detection, fluency illusion). The supporting evidence column links each principle to the findings' points (e.g., F20, F21 for delayed feedback). Together, the principles operationalize the shift from a GPS that commands to a compass that orients, preserving learner agency while building durable metacognitive skills.

Table 2. Five Design Principles for Metacognitive AI

Principle	Description	Metacognitive Phase Supported	Primary Risk Addressed	Key Supporting Evidence
1. Ask before you act	Prompt the learner to articulate a plan before AI helps	Planning	Bypassed strategy selection; passive reception	Azevedo & Gaševic (2019); Chi (2000)
2. Delay and fade feedback	Provide graduated hints rather than full solutions; reduce support as competence grows	Monitoring	Suppression of error detection; learned helplessness	Butler & Winne (1995); Renkl & Atkinson (2003)
3. Mandatory reflection pauses	Insert structured post-task prompts (e.g., "What was confusing?")	Evaluation	Skipped consolidation; failure to update strategic knowledge	Wise & Hsiao (2019); Roelle et al. (2017)
4. Make the process visible	Reveal AI's reasoning path, decision points, and alternatives; invite comparison	Monitoring & Evaluation	Fluency illusion; shallow understanding	Rozenblit & Keil (2002); VanLehn (2011)
5. Calibrate confidence	Ask learners to rate their confidence before revealing answers; track calibration over time	Monitoring (self-assessment)	Overconfidence; illusion of explanatory depth	Dunlosky & Rawson (2012); Hacker et al. (2008)

Synthesis of Cross-Domain Evidence (formerly Literature Review)

The analysis confirms that habitual GPS use provides a validated model of cognitive disuse atrophy (Dahmani & Bohbot, 2017; Ishikawa et al., 2008). Metacognition, the internal cognitive navigation system, is highly susceptible to offloading (Flavell, 1979; Hattie, 2009). Early AI-in-education research shows that unreflective use of generative AI can degrade independent performance and metacognitive accuracy (Bastani et al., 2025; Prather et al., 2024). The principle of prompting learners to generate plans and explanations aligns with broader generative learning strategies shown to enhance retention and transfer (Fiorella & Mayer, 2016). The structural parallels between GPS and AI offloading are striking: performance during tool use is maintained; degradation becomes visible only when the tool is removed; the mechanism is active adaptation (Risko & Gilbert, 2016); and outcomes are contingent on design.

Implications for Practice, Policy, and Future Research

For EdTech developers: The five principles provide a concrete design checklist, prompting, fading, pausing, process visibility, and calibration are technically implementable. The industry must shift from optimizing seamlessness to optimizing durable learning. Public commitments to learning efficacy standards (analogous to nutritional labeling) could align user expectations with educational goals.

For educators: The compass metaphor offers a heuristic for selecting AI tools: “Does this tool make student thinking visible, or does it hide cognitive work?” Professional development should equip educators to evaluate AI tools on metacognitive affordances, not only content accuracy. Emerging work on teacher-AI co-design suggests that involving educators in tool configuration can better align AI support with classroom metacognitive goals (Holstein et al., 2019). Teachers can also adopt the five principles independently, structuring activities that interleave AI assistance with student-generated planning and reflection.

For policymakers: Regulatory discussions focus on data privacy, algorithmic bias, and content accuracy, but cognitive consequences are unaddressed. Educational AI should be subject to cognitive efficacy standards analogous to medical devices: any tool that systematically offloads a core cognitive function must demonstrate that it does not produce long-term dependence or atrophy in that capacity. Zimmerman's (2018) model of self-regulated learning emphasizes cyclic phases of forethought, performance, and reflection, all of which can be short-circuited by AI that provides immediate solutions.

Future research priorities: (1) Longitudinal studies tracking metacognitive development in AI-using cohorts over months/years; (2) Neuroimaging studies to investigate whether cognitive offloading to AI produces measurable changes in prefrontal/ hippocampal function; (3) Classroom-based design experiments testing the five principles as an integrated package; (4) Research on individual differences (prior knowledge, executive function, age) to inform adaptive systems. Molenaar's (2022) concept of hybrid human-AI regulation provides a useful framework for understanding how shared control between learner and AI can preserve metacognitive engagement.

Limitations: This integrative review is inherently interpretive; the GPS-AI analogy is functional, not neuroanatomical. The AI-in-education evidence base remains small and predominantly correlational. The five principles are theoretically grounded but empirically untested as an integrated set. Generalisability across AI tools, populations, and cultural contexts is unknown. Unaddressed dimensions include motivation, creativity, epistemic trust, and sociotechnical issues (bias, privacy, political economy).

Recommendations: Longitudinal and experimental studies are urgently needed. Neural and cognitive process measures should be employed. Developmental and individual differences research is essential. Design-based implementation research in authentic classroom settings will refine the principles. For educators, a metacognitive lens for tool selection is recommended. For developers, the five principles should be the default features. For policymakers, cognitive efficacy standards should complement existing regulations.

CONCLUSION

This paper opened with a student, Lena, who had arrived at a correct physics answer but could not explain her own thinking. Her situation was not a failure of intelligence or effort, but a symptom of a cognitive design problem: the AI tool she used had navigated the problem space so completely on her behalf that her inner compass had been left unused. Through an interdisciplinary synthesis of cognitive neuroscience, educational psychology, and early AI in education research, we have argued that Lena's experience is not an isolated anecdote but a predictable consequence of unexamined cognitive offloading and, crucially, that it is a consequence we can design against.

The central argument rests on three pillars: (1) habitual GPS reliance produces cognitive disuse atrophy (Dahmani & Bohbot, 2017; Ishikawa et al., 2008; Maguire et al., 2006); (2) metacognition is a powerful predictor of achievement that requires deliberate practice (Flavell, 1979; Hattie, 2009; Azevedo & Gaševic, 2019); (3) current AI tools can bypass the metacognitive loop, producing immediate gains that mask impaired independent capability (Bastani et al., 2025; Prather et al., 2024; Roelle et al., 2017). By threading these strands together through the metaphor of AI as a "GPS for thinking," we have named the risk and pointed toward a remedy: redesign AI partners as compasses that provide orientation, demand engagement, and strengthen the learner's internal sense of direction.

From this reconceptualization, we derived five evidence-informed design principles: prompt planning before assistance; delay and fade feedback; embed mandatory reflection pauses; make AI reasoning visible; and calibrate learner confidence. These principles are neither technically prohibitive nor pedagogically radical; they build upon decades of research on self-explanation (Chi, 2000), feedback timing (Butler & Winne, 1995), desirable difficulties (Bjork & Bjork, 2011), and metacognitive scaffolding (Azevedo & Gašević, 2019; Wise & Hsiao, 2019).

Limitations notwithstanding, AI is already in classrooms. Inaction is a design choice, one that defaults to short-term task completion at potential long-term cost. The recommendations advanced here for researchers, developers, educators, and policymakers are a starting point for coordinated action. In a world increasingly saturated with cognitive prosthetics, the most important educational outcome may be the capacity to navigate the world of ideas with an internal compass that remains, even when the screens go dark, unmistakably our own. This capacity is the foundation of lifelong learning, critical thinking, and intellectual agency. Losing it to unexamined convenience would be a quiet catastrophe.

We end where we began, with a student. Not Lena frozen at the question, but a different student a year later, in the same classroom, using the same AI, this time redesigned. She types a problem, and the AI asks her to sketch her approach first. She wrestles with a hint, recalibrates her confidence, and, after the task, pauses to reflect. When her teacher kneels beside her and asks how she solved it, she can say: I started here, I got stuck there, I tried this, and I know what I would do differently next time. The AI has helped her, but it has not replaced her. It has been a compass, not a GPS. Moreover, the map she carries now is her own.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this paper.

REFERENCES

Aleven, V. A. W. M. M., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26*(2), 147179.

Azevedo, R., & Gaševic, D. (2019). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. *Computers in Human Behavior, 96*, 207210.

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics—proceedings *of the National Academy of Sciences, 122*(26), e2422633122.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays* *illustrating fundamental contributions to society* (pp. 5664). Worth Publishers.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245281.

Chen, P., & Chavez, O. (2022). Metacognitive scaffolding in intelligent tutoring systems: A meta-analysis. *Computers & Education, 185*, 104529.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 5, pp. 161238). Lawrence Erlbaum.

Dahmani, L., & Bohbot, V. D. (2017). Habitual use of GPS negatively impacts spatial memory during self-guided navigation. *Scientific Reports, 7*, 41128.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271280.

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28*(4), 717741.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906911.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*(1), 4556.

Gramann, K., Hoepner, P., & Karrer-Gauss, K. (2017). Modified navigation instructions for spatial navigation assistance systems lead to incidental spatial learning. *Frontiers in Psychology, 8*, 193.

Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition and Learning, 3*(2), 101121.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool for teacher-AI collaboration. *International Journal of Artificial Intelligence in Education, 29*(3), 350385.

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M. P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M. C., & Vedel, I. (2018). Mixed Methods Appraisal Tool (MMAT) version 2018. *Education for Information, 34*(4), 285291.

Ishikawa, T., Fujiwara, H., Imai, O., & Okabe, A. (2008). Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. *Journal of Environmental Psychology, 28*(1), 7482.

Kirschner, P. A., & Hendrick, C. (2020). *How learning happens: Seminal works in educational psychology and what they mean in practice*. Routledge.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.

Maguire, E. A., Woollett, K., & Spiers, H. J. (2006). London taxi drivers and bus drivers: A structural MRI and neuropsychological analysis. *Hippocampus, 16*(12), 10911101.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097.

Molenaar, I. (2022). The concept of hybrid human-AI regulation: Exemplifying how to support young learners' self-regulated learning. *Computers and Education: Artificial Intelligence, 3*, 100057.

Mollick, E. R., & Mollick, L. (2023). Using AI to implement effective teaching strategies in classrooms: Five strategies, including prompts. *SSRN Electronic Journal*.

OECD. (2021). *AI in education: Promises and pitfalls*. OECD Publishing.

Pozuelos, J. P., & Moreno, R. (2021). Neural correlates of metacognitive monitoring: A meta-analysis of functional neuroimaging studies. *Cognitive Neuroscience, 12*(2), 89101.

Prather, J., Reeves, B. N., Denny, P., Becker, B. A., Leinonen, J., Luxton-Reilly, A., Powell, G., Finnie-Ansley, J., & Savelka, J. (2024). The robots are coming: On the potential of ChatGPT for computing education. In *Proceedings of the 2024 Innovation and Technology in Computer Science Education Conference (ITiCSE 2024)* (Vol. 1, pp. 112118). ACM.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*(1), 1522.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676688.

Roelle, J., Schmidt, E. M., Buchau, A., & Berthold, K. (2017). Effects of informing learners about the dangers of cognitive offloading on their offloading behavior and achievement. *Journal of Educational Psychology, 109*(7), 971987.

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 26*(2), 582599.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*(5), 521562.

Schmidt, R. A., & Bjork, R. A. (2017). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. In *The psychology of learning and motivation*, (67), 145. Academic Press.

Schrader, C. (2020). How to support metacognition in digital learning environments: A review of recent advances. *Journal of Educational Computing Research, 58*(5), 9871012.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*(4), 460475.

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*(4), 351371.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333339.

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology, 8*, 45.

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review, 4*(3), 356367.

Tyndall, J. (2010). AACODS checklist for appraising grey literature. Flinders University.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197221.

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*(1), 314.

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing, 52*(5), 546553.

Wiley, J., & Jensen, J. (2020). The illusion of explanatory depth in digital learning: Why students overestimate their understanding. *Educational Psychology Review, 32*(3), 721744.

Wise, A. F., & Hsiao, Y. T. (2019). Self-regulation in online discussions: Aligning data streams to model and support normative behaviors. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK19)* (pp. 220229). ACM.

Zheng, L., Li, X., & Chen, F. (2019). Effects of metacognitive prompts on learning outcomes in e-learning environments: A meta-analysis. *Journal of Computer Assisted Learning, 35*(2), 179191.

Zimmerman, B. J. (2018). Developing self-regulating learners: The role of feedback and self-reflection. *Educational Psychologist, 53*(3), 153168.