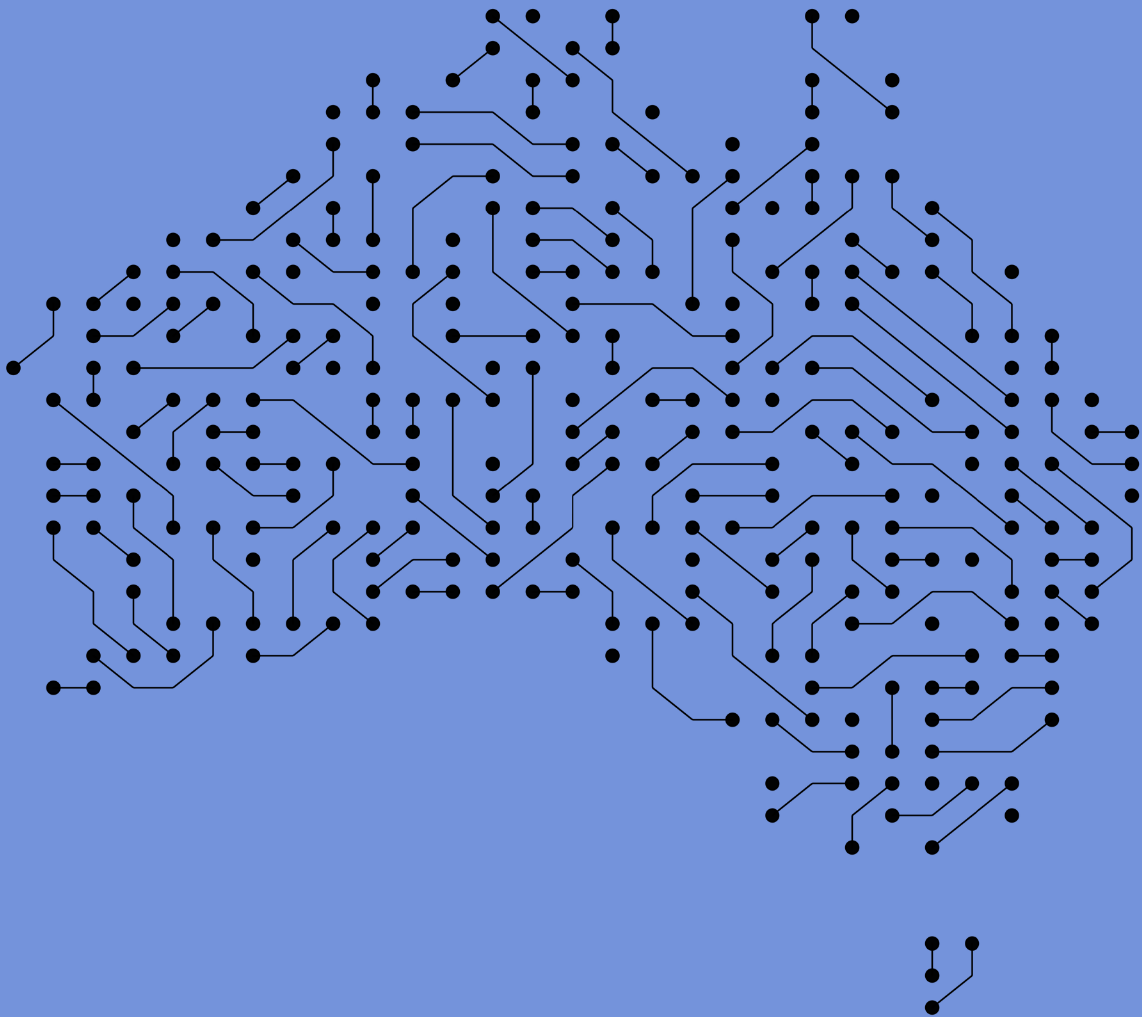


Artificial intelligence, cognitive offloading and implications for education

March 2026

Prof Jason M. Lodge and Prof Leslie Loble AM

UTS Centre for Social Justice and Inclusion

Network for Quality Digital Education

This report forms part of the work program supporting the Australian Network for Quality Digital Education. The Network brings together leaders from across education, industry, social purpose and philanthropic organisations, government and research, in the common purpose of ensuring that all Australian students benefit from the best educational technology (edtech), and the benefits of edtech are leveraged to tackle the persistent learning divide. Members of the Network have provided valuable engagement, input and feedback as part of the report's development, though the report does not represent a consensus or endorsed Network view.

The Network is chaired by Leslie Loble AM, who is Industry Professor at the UTS Centre for Social Justice and Inclusion (socialjustice.uts.edu.au).

Professor Jason M. Lodge is Professor of Educational Psychology and Director of the Learning, Instruction, and Technology Lab in the School of Education, The University of Queensland.

How to cite this report

Lodge J. M. and Loble L (2026). *Artificial intelligence, cognitive offloading and implications for education*, University of Technology Sydney, doi:10.71741/4pyxmbnjaq.31302475.

Copyright information

Any enquiries about or comment on this publication should be directed to edtechnetwork@uts.edu.au.

Accessibility

This report is partially conformant with WCAG2.1 level AA. Known limitations and alternatives are listed below:

- Image descriptions — images and graphs do not have long image descriptions because we have retrofitted accessibility requirements onto an existing template and layout.
- Main body text does not meet requirement of 12-point size because we have retrofitted accessibility requirements onto an existing template and layout. Alternatives are to zoom in or use a screen reader.

Acknowledgements

The authors are grateful that the Australian Network for Quality Digital Education is hosted by the

UTS Centre for Social Justice and Inclusion, and supported by the Paul Ramsay Foundation. The authors appreciate the valuable expertise and guidance of the Network members and others, which has enhanced the report.

We acknowledge the Traditional Owners of Country throughout Australia and pay our respects to Elders past and present.

Paul Ramsay Foundation

262 Liverpool St, Darlinghurst NSW 2010
paulramsayfoundation.org.au

PRF is a philanthropic foundation. The late Paul Ramsay AO established the Foundation in his name in 2006 and, after his death in 2014, left most of his estate to continue his philanthropy for generations to come.

At PRF, we work for a future where people and places have what they need to thrive. With organisations and communities, we invest in, build, and influence the conditions needed to stop disadvantage in Australia.

This research was funded by PRF (grant number 5040). Any opinions, findings, or conclusions expressed in this report are those of the authors and do not necessarily reflect the views of the Foundation.

Table of Contents

This report investigates a profound new challenge driven by AI's power to rapidly access information and provide a semblance of thinking: the risk that students will outsource too much of the cognitive work that is crucial to establishing the knowledge, skill and 'thinking infrastructure' that enables both schooling success and lifelong capacity for ongoing learning, understanding, reflection, creativity and achievement.

Preface

Artificial intelligence (AI) and especially generative AI are propelling a new dynamic for Australian education, simultaneously unlocking compelling opportunities and substantial challenges for teaching and learning. Teachers and students find themselves on the front line of this paradox as they navigate in real time questions of how to best use AI for learning and knowledge gain.

Today, in Australia, nearly 80 percent of students report using artificial intelligence (Thomas et al. 2025), and two-thirds of early secondary teachers — fourth highest usage in the OECD (OECD 2025) — merely three years since ChatGPT burst through.

As AI becomes a near-universal feature of Australian education, we cannot disentangle discussions of AI from discussions of what will make Australian education most effective and equitable. Yet the education sector often feels like it's the tail while others have the whip hand. Just as it gets on top of one aspect of AI, significant new dimensions emerge.

For teachers and for policy makers, all of this can sometimes feel overwhelming, especially when there is such limited research or even consistent experience with a technology that is intentionally designed to keep changing, from prompt to prompt, version to version, year to year. Retaining agency over the design, use and governance of AI thus becomes an essential component of successful integration of AI in education.

The AI dynamic is nuanced and complex in education. It can both counteract — or compound — non-technological factors that propel learning gaps and educational outcomes, from teacher shortages to uneven distribution of resources and concentrations of disadvantage. The impact depends on decisions regarding an AI tool's quality, accessibility and, above all, effective pedagogical use.

On the positive side of the ledger, some of the strongest available evidence points to sustained learning gains from AI-enabled adaptive tutoring systems (Loble & Hawcroft 2022). Education technology (edtech) also can assist students with disability (OECD 2026), who now comprise a quarter of Australian school classrooms (ACARA 2024). AI reduces teacher workloads, freeing time for more valuable educational interactions (NSW 2024). And by providing quality teaching resources, well-designed digital tools can help support consistent access to high-quality, content-rich curriculum, a key factor in securing educational equity (Jensen et al. 2023).

Securing the benefits of this technology relies on access to quality resources, digital inclusion, skills and understanding, and teacher expertise in how to effectively incorporate AI for maximal student learning gain. Without these foundational elements, the risk rises both for uneven distribution of AI's benefits and for challenges in obtaining positive educational outcomes (Loble & Stephens 2024).

This report investigates a related, and profound, new challenge that we face with AI's power to rapidly access information and provide a semblance of thinking: the risk that students will outsource too much of the cognitive work that is crucial to establishing the knowledge, skill and 'thinking infrastructure' that enables both schooling success and lifelong capacity for ongoing learning, understanding, reflection, creativity and achievement.

There is a growing body of evidence that using AI can short-circuit the cognitive effort required for sustainable, deep learning, thus creating "false mastery" with potentially long-term consequences (OECD 2026). This cognitive offloading from human to AI is especially risky for school students ('novice' learners who are building foundational knowledge and skills) when they turn to AI as a tempting substitute, not an amplifier, increase their dependency on the tool and lose access to deeper learning.

Cognitive outsourcing also introduces extra equity risks. Research suggests that students who possess high levels of content knowledge and strong metacognitive skills are better able to leverage AI to accelerate and deepen their learning and critical thinking (Hong et al. 2025). Conversely, students lacking such skills, **often those already experiencing disadvantage, are potentially more susceptible to harmful offloading and missing the learning they need.** The unstructured use of AI risks even wider equity divides (Loble & Stephens 2024).

The good news is that research studies also suggest these harmful effects can be counteracted through purposeful teaching and learning strategies and effective design of AI education technology. These strategies reinforce the importance of quality teaching, with AI in a subsidiary, supporting position.

So, while the extent and scale to which students shift their knowledge and skill acquisition to AI raises fundamental questions for teaching and learning — questions that cannot be answered solely by teaching AI literacy — it also brings an important opportunity to validate and bolster the role of teachers.

This does not mean tackling harmful cognitive outsourcing is solely the responsibility of teachers, however. The ecosystem that supports effective teaching and learning (including quality curriculum) must respond, too, and how we navigate the nuance of positive and negative cognitive outsourcing with AI will depend on good decisions across schooling and public policy.

Part of empowering teachers also means ensuring we direct the design of AI tools so teachers have trustworthy resources to use. A large proportion of educational technology contains little explicit learning content, nor grounding in evidence-based pedagogy (UNESCO 2023), especially general-purpose AI chatbots lacking educational guidance or reliable evidence basis. The work tasked by Education Ministers for development of edtech national standards and quality assurance procedures is essential and urgent.

Responding to the challenge

The Australian Network for Quality Digital Education has brought together leaders across education, government, industry, social purpose, research and philanthropy committed to harnessing edtech's potential to improve learning outcomes, especially for those students experiencing disadvantage, while countering risks that undermine positive impact.

At a recent forum, the Network considered the challenge of AI and harmful cognitive offloading. **Two critical leverage points** emerged:

1. strategies to help teachers most effectively deploy AI, drawing on the substantial evidence of what already is known to work best in teaching and learning and by explicitly structuring and scaffolding the student use of AI; and
2. educational design of AI tools so that they amplify teacher expertise to build, not relinquish, the student cognitive effort required for lasting knowledge.

This report outlines specific promising pathways, backed by evidence, that align with these priorities, including:

+ using explicit teaching strategies that help students to offload lower-order tasks while building self-regulated learning capability and critical thinking
+ bringing clear metacognitive prompts into the learning process to encourage deeper inquiry and reflection and help students become self-regulating and motivated learners
+ teaching explicitly and within domain knowledge the critical thinking capabilities that will help students understand, evaluate and consider complex content
+ ensuring teachers retain the necessary agency to help shape and select AI tools that will support effective pedagogical approaches
+ accelerating adoption of the draft national education design standards for AI tools, to ensure they are designed to bolster students' learning effort, mastery and cognitive agency

The potential outsourcing of learning to AI introduces high stakes for students' successful attainment of essential bedrock knowledge, capabilities and cognitive processes (see Oakley et al. 2025). Two decades ago, there was early evidence of digital technologies' cognitive impact (Fogg, B.J. 2003), a precursor of today's concerns and Australian action on social media. Now, there is increasingly strong evidence that AI used in schooling must also be closely governed and directed, not only for safety but to ensure we have confidence in the AI tools themselves and that teachers are well-supported with effective pedagogical strategies. This report outlines that while AI may be a new technological vector for education, the strategies for its successful integration require a strong pedagogical response: enhancing the central role of teachers; drawing upon well-researched approaches for quality teaching and learning; and ensuring close attention to equity.

While AI may be a new technological vector for education, the strategies for its successful integration require a strong pedagogical response: enhancing the central role of teachers; drawing upon well-researched approaches for quality teaching and learning; and ensuring close attention to equity.

Executive summary

The rapid adoption of AI (particularly generative AI) presents a novel challenge to K-12 education. It has the capacity to function as an interactive cognitive partner, bringing the concept of cognitive offloading (outsourcing mental work) to the forefront.

This report analyses this phenomenon through the lens of human cognitive architecture (Cognitive Load Theory), framing the central problem as a conflict: the capacity of AI to bypass the cognitive effort required to build the deep, long-term knowledge that underpins expertise and critical thinking.

The report's central finding is a critical distinction between two forms of offloading:

+ **Beneficial offloading** occurs when AI is used to manage extraneous cognitive load (e.g., checking grammar), freeing a learner's limited working memory to focus on essential, intrinsic tasks.
+ **Detrimental offloading** (outsourcing) occurs when a learner uses AI to bypass this intrinsic cognitive effort (the desirable difficulties) required to build long-term knowledge schemas. This offloading also seems to extend to vital metacognitive and self-regulated learning capabilities, compounding the negative impact of outsourcing on learning.

Emerging data support the observation that unstructured AI use trends toward detrimental offloading, creating a performance paradox: students' short-term *performance* on tasks improves, while their durable, long-term *learning* is harmed. This trend appears to be driven by the fluency of AI-generated output, which creates an illusion of competence and encourages metacognitive laziness, leading learners to abdicate the generative effort required to build deep knowledge.

The impact of AI is not primarily technologically deterministic; it is pedagogical. While unstructured use risks cognitive atrophy, the report finds that pedagogically structured interventions, such as explicit teaching, Load Reduction Instruction (LRI), and integrated metacognitive prompts, can successfully foster the self-regulated learning, critical thinking and the deep engagement required for learning.

Glossary of key terms

Artificial intelligence	A broad term that refers to a range of emerging and evolving technologies. While acknowledging this diversity, the report uses "AI" as a collective term to refer primarily to generative AI (like large language models) and automated decision-making systems (which are sometimes referred to as 'AI agents' or 'agentic AI').
AI agents (or 'agentic AI')	Automated decision-making systems.
Beneficial cognitive offloading	Occurs when a learner uses a tool (like AI) to outsource **extraneous load** (e.g., checking grammar). This action frees up limited working memory resources to focus on the **intrinsic** work of learning (e.g., structuring an argument, synthesising sources).
Cognitive Load Theory (CLT)	An instructional theory based on the structure of human memory. It provides a framework for designing teaching that respects the severe limits of **working memory** to optimise the construction of **schemas** in **long-term memory**.
Cognitive mirror	A pedagogical strategy where an AI is designed to act as a teachable novice. The AI feigns confusion and asks clarifying questions, which forces the human learner to engage in the effortful act of explanation and reflection (known as the **Protégé effect**).
Cognitive offloading	The formal definition is the use of physical action to alter the information processing requirements of a task to reduce cognitive demand. It means outsourcing mental work to an external resource. While writing a to-do list is a simple example, AI allows for offloading complex tasks like analysis, synthesis, and creation.
Cognitive partner	A term used to describe AI as an interactive, fluent, and responsive tool. Unlike a simple calculator, an AI partner can be prompted to perform the very cognitive tasks (like summarising, analysing, and creating) that are traditionally associated with the process of learning.
Desirable difficulties	The concept that durable, long-term learning is not meant to be effortless and requires a degree of cognitive effort or challenge. The risk of AI is that it may encourage students to bypass this essential cognitive struggle, which harms their long-term learning.
Detrimental cognitive offloading (outsourcing)	Occurs when a learner outsources the **intrinsic cognitive work itself**. This is an attempt to bypass the work of learning. An example is asking AI to "write me an essay," which bypasses the entire schema-construction process (generation, retrieval, analysis, synthesis).

Domain-specific knowledge	The deep, well-organised foundation of knowledge within a specific field (e.g., a scientist's knowledge of experimental design or a historian's knowledge of social context). The paper argues that this knowledge, stored in **long-term memory**, is the necessary foundation for critical thinking in most cases.
Evaluative judgement	The capability to make decisions about the quality of work of oneself and others. As defined by Tai et al. (2017), developing this capability is a proposed goal of education to enable students to improve their work and meet future learning needs. It involves two key components: understanding what constitutes quality and applying this understanding to appraise work.
Extraneous load	A concept from **Cognitive Load Theory**, this is the unnecessary or non-essential cognitive load imposed by *how* information is presented, or the activities learners are asked to do (e.g., confusing instructions, distracting layouts). Effective teaching aims to minimise this load.
Fluency (in learning)	Refers to the ease of processing learning materials, such as watching a high-quality video. AI is described as "fluency on demand" because its output is coherent, confident, and articulate. This is dangerous as it can create an **illusion of competence**, where learners mistake the ease of processing for the depth of learning.
Generation effect	A cognitive principle demonstrating that when students are forced to *generate* an answer from a cue (a **desirable difficulty**), they have significantly better long-term vocabulary retention than students who simply review the word passively.
Human cognitive architecture	The underlying structure of the mind's information-processing systems. For the purposes of teaching, it is understood to have two key interacting components: **working memory** and **long-term memory**.
Illusion of competence	A cognitive bias where learners greatly overestimate how much they have learned. The **fluency** of AI-generated text or multimedia can create this illusion, acting as a misleading cue that signals to the learner that deep cognitive engagement is no longer necessary.
Intrinsic load	A concept from **Cognitive Load Theory**, this is the inherent, unavoidable complexity of the learning material itself. This is the "good" load associated with the core concepts and the effortful process of building schemas (that can, nonetheless, also lead to overload if there is too much).

Learn the concept goal	A learner's motivation that is contrasted with a **task completion goal**. Learners with this goal are more likely to *actively modify* AI-generated text, which leads to significant improvements in their work.
Load Reduction Instruction (LRI)	A pedagogical framework that adapts explicit instruction principles that was not specifically developed for but has utility for AI. It involves using AI to provide scaffolding, structured practice, and feedback to manage the cognitive burden on the learner and enable progressive independence.
Long-term memory	A core component of human cognitive architecture. It is the vast, organised store of all acquired knowledge, experiences, and procedures, and its capacity is considered unlimited. Learning is defined as the process of integrating new information into long-term memory by building schemas.
Matthew Effect (with AI)	A term used to describe a worrying metacognitive equity gap in an AI-mediated world. Students who *already* have high domain knowledge and strong metacognition will use AI to accelerate their learning. In contrast, students *without* these skills will fall prey to detrimental offloading and fall further behind.
Metacognition	The process of thinking about one's own thinking. This includes essential self-regulated learning processes like planning, monitoring, and revision.
Metacognitive equity gap	A significant and novel equity risk identified in this report. This gap arises because the cognitive risks of AI (like metacognitive laziness and the illusion of competence) disproportionately affect *novices* and those with weaker self-regulation skills, thus potentially widening existing equity divides.
Metacognitive laziness	A term coined by Fan et al. (2024) to describe how the convenience of AI can undermine learners' engagement in essential self-regulatory processes. The learner, in effect, abdicates their metacognitive responsibilities to the tool, depriving themselves of the opportunity to develop these skills.
Metacognitive prompts	A pedagogical intervention, often integrated into an AI environment, designed to counter metacognitive laziness by explicitly demanding and scaffolding a learner's self-regulated learning. These prompts are designed to make users pause, reflect, and assess their understanding, leading to more active engagement, enhanced self-regulated learning, and improved metacognitive knowledge.
Outsourcing	A commonly used term describing detrimental offloading.

Performance paradox	A performance paradox is identified where AI can boost a student's performance on an immediate task while simultaneously diminishing the durable learning that is the goal of education. For example, students using AI may solve problems effectively, but their learning suffers once the AI is taken away because the scaffolded performance did not translate into durable knowledge.
Protégé effect	A learning benefit that occurs when a person is forced into the effortful, generative act of explanation and reflection. This is the mechanism triggered by the cognitive mirror pedagogy.
Schemas	Complex knowledge structures in long-term memory that organise information according to how it will be used. Learning is the process of building these schemas. An expert is someone who has built vast and complex schemas in their long-term memory and knows when and how to apply them.
Self-regulated learning (SRL)	The ability to manage one's own learning. This involves processes like planning, monitoring, and revision. The report notes that engaging in SRL *itself* creates a cognitive load, which is why students, seeking efficiency, may be tempted to offload these processes to AI.
Task completion goal	A learner's motivation, often driven by a desire for efficiency, that contrasts with a learn the concept goal. Learners with this goal are more likely to passively accept AI-generated text, which can lead to a *decrease* in the quality of their work.
Verification partner	A pedagogical model that reframes AI's role away from being an answer oracle. In this model, the human learner maintains primary cognitive agency and is responsible for continuously evaluating and correcting the AI's output, thereby adopting a verification mindset.
Working memory	A core component of human cognitive architecture. It is the conscious part of the mind that actively processes new information (e.g., holding a phone number while dialling). Its defining feature is that it is severely limited and can only process a minimal number of new information elements at one time. Exceeding this limit causes cognitive overload, which inhibits learning.

> The true educational risk of AI is not simply that students will use it to cheat on an essay. The far more profound risk is that AI may fundamentally interfere with the cognitive processes of knowledge construction and verification, the very processes that build the long term memory stores and subsequent skills upon which the majority of critical thinking depends.

Introduction

Even before the global impact of the COVID-19 pandemic on education, a clear trend was emerging across education systems worldwide toward the increased use of digital learning environments and tools.

These trends extend back more than 30 years to when the first digital technologies systematically impacted education. The forced move to online learning during the pandemic accelerated this trend, demonstrating what was possible, even with little preparation. However, as with those technologies, the flexibility and engagement granted by new digital tools often come with a cognitive cost (Lodge 2023). That cost is becoming progressively evident as artificial intelligence (AI) increasingly impacts on learning and teaching (Horvath 2026).

The public launch of ChatGPT in late 2022 has triggered a new, more profound acceleration, moving the debate beyond blended learning to the very foundations of human cognition. These technologies have evolved and continue to advance. Beginning with large language models and generative AI, automated decision-making systems (sometimes referred to as 'AI agents' or 'agentic AI') have also come to the fore. For the purpose of this report, these technologies will be referred to collectively as 'AI'. That is not to undersell the reality that there is a wide range of emerging and evolving technologies here that require attention.

These new technologies have, perhaps unsurprisingly, led to a polarised debate, often dominated by "booster" narratives that frame AI (particularly, generative AI) as a saviour for education, and "doomster" narratives that see it as the end of original thought. Indeed, the stakes are high. Prominent scholars have warned that humankind may be at "peak humanity," with AI posing fundamental risks to cognitive development, agency, and a future of "de-education" (Hamilton et al. 2023).

The purpose of this report is not to engage in that debate, but to move beyond it. It seeks to provide a robust but accessible analysis of the novel interaction between AI and the established mechanisms of human learning. It is recognised here that the research in this area is still nascent. Much more is yet to be done to understand the impacts of AI on learning. AI is not a simple tool like a calculator or a search engine. It can function as an interactive, fluent, and responsive *cognitive partner* (Lodge et al. 2023). AI tools can be prompted to summarise, analyse, create, critique, and explain. In doing so, **these technologies offer to perform (or at least simulate) the cognitive tasks (generation, retrieval, synthesis, and evaluation) that we have traditionally associated with the process of learning**.

This capability brings the psychological concept of cognitive *offloading* (Risko & Gilbert 2016), the outsourcing of mental work to an external resource, from a niche academic interest to a central, practical challenge for every teacher, educational leader, and policymaker. What happens to learning when a student has a partner capable of providing a fluent, coherent, and often (but not always) correct answer, thereby allowing the student to bypass the effortful cognitive work required to generate that answer themselves? Does this offloading free the student for higher-order thinking, or does it prevent the very knowledge construction required for that thinking to occur?

One way to conceptualise how dramatic a shift learning and teaching are experiencing due to AI is to consider what it means for students to have a partner like AI in their learning. Every task or learning activity is essentially now a group activity. It just so happens that the other member or members of the group are machines that have practically all human knowledge at their fingertips (in their databases/algorithmic weights). Like any other group activity, students can benefit from that collaboration or get the smart kid to do all the work for them.

This report aims to provide a foundational cognitive framework to address these questions based on the current state of the research literature. It will provide an outline of human cognitive architecture, grounded in Cognitive Load Theory (Kalyuga & Plass 2025; Sweller et al. 2011), to explore the relationships among AI, cognitive offloading, and their implications for K–12 teaching and learning. With the acknowledged limitations of AI, such as its capacity to “hallucinate” and replicate societal biases (Yan et al. 2024), the report will pay specific attention to the implications for education’s most vital and enduring goals: the development of expertise, mastery, critical thinking and self-regulated learning.

The enduring primacy of knowledge

For decades, educational discourse has been animated by a perceived tension between knowledge and skills. Proponents of "21st-century skills" have often argued that in an age of rapid change and instant information access, the ability to *find* information (a skill) is more important than the ability to *know* it (a knowledge store).

This perspective posits that skills like critical thinking, creativity, and collaboration are generic, transferable competencies that can be taught in the abstract.

Insights from cognitive science, however, challenge this assumption. The evidence overwhelmingly indicates that high-order skills, particularly critical thinking, are not generic (Willingham 2019; Tricot & Sweller 2014). Instead, they are commonly deeply intertwined with and dependent upon a well-organised foundation of domain-specific knowledge stored in long-term memory. As Willingham (2019) argues, "Thought processes are intertwined with what is being thought about." It is not possible to engage in critical thinking when one has nothing to think critically about. A person does not simply think critically in a vacuum. A scientist thinks critically about a flawed methodology by drawing on a vast store of knowledge about experimental design. A historian thinks critically about a primary source by drawing on their knowledge of the document's social, political, and historical context (Willingham 2019).

The deep structure of a problem is visible only to those with the requisite domain knowledge; novices, by contrast, are distracted by the problem's "surface structure" and fail to see the underlying principles (Willingham 2019). This view is powerfully reinforced by Tricot and Sweller (2014), who distinguish between "biologically primary" generic skills (which have evolved and are largely unteachable, like speaking) and "biologically secondary" cultural knowledge (which is the entire purpose of schooling and *requires* explicit instruction), based on the work of Geary (1995). Tricot and Sweller argue that "domain-specific knowledge held in long-term memory... [is] arguably the most important factor, and possibly the only factor, determining acquired cognitive performance" (p. 265).

In this framework, **you cannot "just Google it" to engage in critical thought. The material retrieved from a search engine exists only in the environment in which it was retrieved. The information has not been integrated into the long-term memory structures (schemas) that are necessary for complex reasoning. This cognitive understanding establishes the central stake of the AI debate.**

However, there is some nuance in this debate that is worth unpacking here as it applies directly to the current debates about the role of AI in learning and thinking. While the long history of work on 21st-century skills shows how difficult it is for skills to transfer, some do. For example, it does not require significant knowledge of a domain to know that one does not have sufficient expertise in it. Recognising and accepting that one is a novice is critical at the beginning of a learning journey. Neither does it require a substantial body of acquired knowledge to understand when one is stuck in their learning and needs to ask for help. Study strategies such as testing oneself have also been shown to enhance learning across contexts (Trumbo et al. 2021). These are established aspects of self-regulated learning (Panadero 2017).

Similarly, some aspects of formal and informal logic, argument analysis, and reasoning skills do seem to transfer (Leibovitch et al. 2024). As Willingham (2019) rightly points out, it is possible and feasible to teach these skills explicitly, which is a topic we will return to later in this report. For the time being, it is important to flag that, while many cognitive skills do indeed require deep domain knowledge, that is not universally true of all cognitive skills.

Notwithstanding the caveat that some skills are transferable, **AI, with its propensity for fluent error and confident "hallucination" (Yan et al. 2024), does not render deep knowledge obsolete; it makes it *more* critical, not less**. A primary critical thinking task in an AI-mediated world becomes one of *verification* or *judgement*. The ability to critically evaluate an AI-produced text, to spot its subtle biases, its logical flaws, or its factual inaccuracies, is almost entirely dependent on the user's own store of domain-specific knowledge. While the term has somewhat fallen out of favour in recent years, these capabilities for interrogating information have previously been referred to as 'information literacy' (Sample 2020).

Achieving the intelligence equilibrium proposed by Grace (2025), in which AI complements rather than replaces human analysis, requires instructional strategies that foster this verification mindset (information literacy). Similarly, Bearman and colleagues (2024) argue that evaluative judgement is the core skill for the age of AI. The true educational risk of AI is therefore not simply that students will use it to cheat on an essay. The far more profound risk is that AI may fundamentally interfere with the *cognitive processes of knowledge construction and verification*, the very processes that build the long-term memory stores and subsequent skills upon which the majority of critical thinking depends.

Human cognitive architecture and the goal-driven learner

To understand how AI impacts learning, we must first have a clear, albeit simplified, model of how learning occurs.

Human cognitive architecture is the underlying structure of the mind's information-processing systems (Sweller et al. 2019). It is important to note here that this is a predominantly rationalist view of mental processes that will not completely capture the complexities of human-machine interaction in learning.

The information-processing paradigm helps explore and understand the mechanisms at work for learners but is limited in its ability to explain the complex dynamics in real-life learning settings. Nonetheless, for the purposes of teaching and learning, this architecture is understood to comprise two key components that interact with one another and are supported by decades of research in psychological science: working memory and long-term memory (CESE 2017).

+ **Working memory:** This is the conscious part of the mind, the system that actively processes new information. It is where you hold a phone number while you dial it or where you mentally combine the elements of a sentence as you read it. The most important feature of working memory is that it is *severely limited* (Baddeley 1986). It is widely held that working memory can only process a very small number of new information elements at one time, perhaps as few as four (CESE 2017). When this limit is exceeded, the learner experiences cognitive overload, which inhibits learning.
+ **Long-term memory:** This is the vast, organised store of all the knowledge, experiences, and procedures a person has acquired. Its capacity is, for all practical purposes, unlimited (CESE 2017). *Learning* is the process of taking new information from the environment, processing it in working memory, and integrating it into long-term memory. This is achieved by building "schemas", complex knowledge structures that organise information according to how it will be used (CESE 2017). An expert is an expert not because they have a better working memory, but because they have built vast and complex schemas in their long-term memory.

Cognitive Load Theory (CLT) is an instructionally relevant theory grounded in this understanding of cognitive architecture (Sweller et al. 2019). It provides a framework for designing instruction that respects the limits of working memory to optimise the construction of schemas in long-term memory.

Recent theoretical reframing of CLT has formalised this model in a way that is essential for understanding the challenges posed by AI for learning (Kalyuga & Plass 2025).

This new work proposes two major modifications:

1. **A two-load model:** It returns to a two-load model, discarding “germane cognitive load” as a separate construct. The effortful process of schema construction is understood as part of *intrinsic load*. Thus, there are only two loads: *intrinsic* (unavoidable) and *extraneous* (non-essential).
2. **A “goal-driven” approach:** This is the most critical modification. Kalyuga and Plass (2025) propose an integrated framework where the *learner’s activities, motivations, and goals*, not just the instructional design, are the central focus of the analysis.

This goal-driven framework provides the key to understanding the challenges AI poses to learning. A learner can use the same tool to pursue two very different goals. Learners, especially novices, are often motivated by a desire for *efficiency* (Zhai et al. 2024), which translates to a task completion goal. Other learners, or the same learner in a different context, may have a ‘learn the concept’ goal. The learner’s goal determines how they interact with the technology.

This modification to CLT goes some way towards recognising the vast range of possible use cases for AI in real learning settings. As has long been argued by Fawns (2022), Dron (2022), and others, technology use in education cannot be represented by a singular, monolithic use case.

In this context, making a broad claim such as ‘ChatGPT is bad for learning’ (e.g. Kosmyna et al. 2025) makes no sense without understanding what the tool was being used for and in what context. A technology like AI is entangled with learning and thinking in numerous, complex ways, driven by many factors, not least the learner’s goals and motivations. Incorporating some aspects of these differences is an important step forward for understanding the nuances associated with sound and poor use of AI in learning and teaching.

Behavioural trace data from learning analytics confirms this complexity and the need to incorporate individual differences, particularly related to goals. In a study of 1,445 AI-assisted writing sessions, learners who *actively modified* AI-generated text (indicating a learn the concept goal) saw significant improvements in their work. Conversely, learners who *passively accepted* the text (indicating a task completion goal) saw a *decrease* in quality (Yang et al. 2025). **The risk of AI is that its convenience and efficiency powerfully incentivise the task completion goal, leading students to frame the *intrinsic* load of learning, the effortful thinking, as something to be outsourced, rather than embraced.** This is particularly the case given that it has been reliably shown that humans find the mental effort associated with learning and thinking to be a negative experience (David et al. 2024).

Table 1: The reframed two-load model of Cognitive Load (adapted from Kalyuga & Plass, 2025)

Load type	Description	Effect on learning	Educator summary and example
Intrinsic load	The inherent complexity of the learning material itself. This is the "good" (unless excessive) load associated with the core concepts and the effortful process of schema construction.	**Necessary**	This is the difficulty of the content. For example, learning to solve a 2-step algebraic equation has a higher intrinsic load than learning a 1-step equation. This load must be *managed*.
Extraneous load	The unnecessary cognitive load imposed by *how* information is presented or the activities learners are asked to perform (e.g., confusing instructions, distracting layout).	**Harmful**	This is the difficulty of the *instruction*. For example, a confusing worksheet layout or a poorly designed interface. Effective teaching aims to *minimise* this load.

Beneficial vs. detrimental cognitive offloading

Cognitive offloading is formally defined as "the use of physical action to alter the information processing requirements of a task so as to reduce cognitive demand" (Risko & Gilbert 2016, p. 676). Writing a to-do list, setting a smartphone reminder, or using a calculator are all acts of cognitive offloading. This is not a new behaviour, but AI makes it possible to offload cognitive tasks of unprecedented complexity, such as analysis, synthesis, and creation. This distinction between AI technologies and previous technologies is critical to how educational institutions and systems need to adapt. It appears that it is now not only possible but very straightforward to offload more than the cognitive aspects of learning.

Crucially, cognitive offloading is not a passive process. It is an active, *metacognitive choice*. Learners, often unconsciously, conduct a cognitive cost-benefit analysis to determine whether it is worth the effort to use their own internal resources (working memory) or to offload the task to an external resource (Dunn & Risko 2016). This offloading can be positive or negative. The impact depends entirely on *what* is being offloaded and what the goal is. This creates the critical distinction at the heart of this report: is the learner offloading an *extraneous* process or an *intrinsic* one?

Beneficial cognitive offloading

Beneficial offloading occurs when a learner offloads *extraneous load* to free up limited working memory resources for the *intrinsic* work of learning. This aligns with the principles of Load Reduction Instruction (LRI), an approach to explicit teaching that aims to manage cognitive burden (Martin et al. 2025). For example, a student writing a complex history essay might use AI to offload the *extraneous* tasks of checking grammar and syntax. This frees their cognitive capacity to focus on the *intrinsic* tasks of evaluating evidence, structuring their argument, and synthesising sources.

This approach has been validated. A 2025 quasi-experimental study by Hong et al. (n=240) provides the empirical support. A group of students was *explicitly taught* to delegate lower-order writing tasks (like brainstorming and grammar) to AI, compelling them to focus their own effort on higher-order analysis and evaluation. This group showed significantly greater improvements in critical thinking than control groups (Hong et al. 2025). This finding demonstrates that **when offloading is intentionally and pedagogically managed, it can successfully support the development of higher-order skills**. This is further supported by Iqbal et al. (2025), who found, in a study of 465 preservice teachers, that the benefits of AI for learning were mediated by *effective* (i.e., lower order) cognitive offloading and *shared* (rather than offloaded) *metacognition*.

Detrimental cognitive offloading (outsourcing)

Detrimental cognitive offloading, or outsourcing, occurs when a learner offloads the *intrinsic cognitive work itself*. As will be discussed further, there is a case for this detrimental offloading also occurring at a metacognitive level. This offloading is an attempt to bypass the 'work of learning'. In this scenario, the student with the task completion goal (as per Kalyuga & Plass 2025) asks AI to "write me an essay about..." By doing so, they are not managing load; they are *bypassing* the entire schema-construction process: the generation of ideas, the retrieval of knowledge, the analysis of connections, and the synthesis of an argument.

The default, unstructured use of AI by students appears to trend toward this detrimental, outsourcing path. The most compelling evidence for this comes from the study of students learning high school mathematics (Bastani et al. 2025). Students who used the AI to get the *answer* to a problem (thereby outsourcing the entire *intrinsic* problem-solving process) suffered harm to their long-term, durable learning.

This is reinforced by a mixed-methods study (n=666) that found a significant negative correlation between the frequency of AI tool usage and critical thinking abilities (Gerlich 2025). Critically, a mediation analysis showed that this negative relationship was *explained* by an increase in cognitive offloading (Gerlich 2025).

Table 2: Examples, beneficial vs. detrimental cognitive offloading

Example K–12 task	Beneficial offloading (process supported)	Detrimental offloading (process outsourced)
Writing an essay	**Task:** "Check my grammar," "Rephrase this sentence for clarity," "Brainstorm some counter-arguments to my thesis." **Offloaded process:** *Extraneous Load* (e.g., syntax, fluency) or *Scaffolding* (e.g., idea generation).	**Task:** “Write me an essay about the causes of World War I,” “Give me the answer.” **Offloaded process:** *Intrinsic Load* (e.g., knowledge retrieval, analysis, synthesis, generation).
	Cognitive impact: Frees working memory to focus on higher-order processes: argument structure, evidence synthesis, and critical analysis.	**Cognitive impact:** Bypasses the *entire* schema-construction process. The learner does not retrieve, analyse, or synthesise information.
	Learning outcome: Stronger critical thinking and writing skills. Students taught this explicit method showed *greater* gains in critical thinking (Hong et al., 2025).	**Learning outcome:** No durable learning; erosion of metacognitive habits (Grace, 2025). The learner’s ability *suffers* once the tool is removed (Bastani et al., 2025).
Maths problem	**Task:** "Explain the quadratic formula in a different way," "Show me a step-by-step worked example," "Generate some practice problems for me." **Offloaded process:** *Extraneous Load* (e.g., confusing textbook explanation) or *Managing Intrinsic Load* (e.g., breaking the problem down).	**Task:** “What is the answer to this problem?” **Offloaded process:** *Intrinsic Load* (e.g., the entire problem-solving process).
	Cognitive impact: Provides adaptive scaffolding to help the learner build a problem-solving schema.	**Cognitive impact:** Bypasses the desirable difficulty of the problem-solving process.
	Learning outcome: Durable schema for solving that problem type.	**Learning outcome:** High *performance* on the task, but *harm* to long-term *learning* (Bastani et al. 2025).

The “performance paradox” in learning with AI

The risk of detrimental offloading is not theoretical. A growing body of empirical research provides clear evidence for a “performance paradox”: AI can boost a student’s performance on an immediate task while simultaneously diminishing the durable learning that is the goal of education (see Yan et al. 2025).

The most critical evidence for this paradox comes from a 2025 study by Bastani et al. on high school mathematics. In a large, randomised experiment involving “nearly 1,000” students, those given access to an AI-powered assistant appeared to facilitate students’ problem-solving while learning. However, their learning suffered once the AI was taken away (Bastani et al. 2025). This finding is the empirical embodiment of the core concern with outsourcing: the scaffolded performance provided by the AI did not translate into durable, independent knowledge (schemas) in the students’ long-term memory.

Other studies reinforce this finding. For example, Darvishi et al. (2024) investigated AI assistance in a peer-feedback context and found a tendency for students to *rely on* rather than *learn from* the AI. When the AI assistance was removed, the students’ performance could not match what they had achieved with the tool. Other studies have found that while AI use leads to cognitive ease and reduced mental effort, this comes at the cost of compromising depth in student inquiry (Stadler et al. 2024). This finding aligns with Kolhatkar’s (2025) findings that cognitive outsourcing reduced independent problem-solving by 35%.

The performance paradox also helps to resolve an apparent contradiction in the research. Several systematic reviews and meta-analyses (e.g. Deng et al. 2024; Han et al. 2025; Wang & Fan 2025) have reported *positive* effects of AI on learning. This “fast science” is critiqued by Weidlich et al. (2025) as an “effect in search of a cause,” suffering from conceptual pitfalls. The chief pitfall is the use of invalid outcome measures. Studies reporting positive effects (e.g., Kestin et al. 2025; Lademann et al. 2025) are often measuring short-term, scaffolded *performance* (the “effect”), such as a post-test or exam, rather than durable, independent *learning* (the “cause”). The performance paradox is the unifying theory: the positive studies are merely measuring the *performance* side of the paradox, while the negative studies, like Bastani et al. (2025), are measuring the *learning* side.

The performance paradox is also aligned with the ongoing discussion since the emergence of ChatGPT that learning is both a process and the outcome of that process (as per Soderstrom & Bjork 2015). Determining the quality of student learning based on the outputs alone is a significant reason why AI has created a fundamental risk to many of the most common assessment tasks assigned to students (Lodge et al. 2023a). The performance paradox is therefore a critical consideration for learning, curriculum and assessment.

Unpacking the paradox: Bypassing "desirable difficulties"

The performance paradox is explained by the cognitive mechanism of desirable difficulties (Bjork & Bjork 2015). Learning is not meant to be effortless. Durable, long-term learning requires a degree of cognitive effort or desirable difficulty (de Bruin et al. 2023). Difficulties and challenges are fundamental to high-quality learning in many contexts (for review, see Lodge et al. 2018). The Start and Stick to Desirable Difficulties (S2D2) framework provides the theoretical basis for *why* students must invest this effort, which AI may encourage students to bypass, to the detriment of their long-term learning (de Bruin et al. 2023).

The temptation to cut corners in learning is not new. Research on academic integrity, mainly in the higher education context (e.g. Ellis & Murdoch 2024), demonstrates that many students have the best of intentions for their studies. However, when put under pressure or when other activities are prioritised, the temptation to outsource becomes too much for many students. This temptation is further exacerbated by the previously mentioned finding that humans generally find the work of learning or thinking to be a negative experience (David et al. 2024).

A 2025 classroom study by Duplice provides a perfect non-AI baseline: the generation effect. Students who were forced to *generate* a vocabulary word from a cue (a desirable difficulty) had significantly better long-term vocabulary retention than those who simply passively reviewed the word (Duplice 2025). This provides a direct, mechanistic explanation for the harm observed by Bastani et al. (2025).

AI, when used as an answer oracle, is the ultimate *passive review* tool. It allows the learner to bypass the generation effect entirely. **By providing the answer, the solution, or the essay, it robs the learner of the very cognitive struggle that is necessary to build lasting knowledge.** For novice learners, this tendency can lead to an exacerbation of the well-known Dunning-Kruger effect, where people tend to vastly overestimate their level of knowledge, particularly in new, unfamiliar domains (Kruger & Dunning 1999).

The link from this detrimental offloading to the erosion of critical thinking has been empirically demonstrated. A 2025 quantitative study by Ejaz et al. (n=350) found a significant *negative* correlation between frequent AI use and critical thinking skills. Furthermore, this research team found a strong *positive* correlation between *higher* cognitive load and *better* critical thinking skills, providing direct empirical support for the desirable difficulties framework and its application to learning with AI (Ejaz et al. 2025).

The lower cognitive load reported by frequent AI users (Ejaz et al. 2025) is not beneficial; it is *detrimental*, as it likely signals that the essential intrinsic processing has been bypassed. This was predicted by pre-2022 research, which found that cognitive offloading was boosting immediate performance but diminishing longer-term learning and memory (Grinschgl et al. 2021).

Metacognitive laziness and the illusion of competence

The most troubling aspect of detrimental offloading is that students engage in it willingly. This reveals that the central problem is one of metacognition and self-regulated learning: the processes of thinking about one's own thinking and managing one's own learning. While this might suggest that the onus is on students to do the work here, **there are pedagogical, technological, and systems factors that can assist students in improving their capacities in these areas**.

The convenience of AI encourages what Fan et al. (2024) have termed "metacognitive laziness." In their randomised study, they found that the convenience of AI can undermine learners' engagement in essential self-regulated learning processes (planning, monitoring, and revision). The learner effectively relinquishes their metacognitive responsibilities to the tool (Fan et al. 2024). The motive for this abdication is explained by Wang and Lajoie (2023): engaging in self-regulated learning (SRL) *itself* creates a cognitive load. Students, driven by a rational desire for *efficiency* (Zhai et al. 2024), choose to bypass this immediate cost of SRL by offloading to AI, thereby failing to achieve the long-term benefit of engaging in self-directed learning.

It has long been argued that it is vital to consider the multiple layers occurring when learners engage with the learning process in a self-directed manner. For example, Nelson and Narens (1994) make the compelling argument that success in learning occurs due to the complex interactions between cognitive and metacognitive components of the process. Lodge and colleagues (2018a) describe the necessary judgement and decision-making processes that occur during and are developed through these interactions. Aspects of this skill base have been referred to elsewhere as 'evaluative judgement' (Tai et al. 2018), which begins to align with notions of critical thinking (Cowan 2010). While there are several overlapping ideas about these metacognitive components of learning, what is nonetheless apparent is that they are essential for the long-term development of expertise.

The poor metacognitive choices that seem to be occurring when learners engage with AI are enabled by a powerful illusion of competence (Lodge 2023). This is a familiar problem in digital learning. Research has long shown that fluent learning materials, such as high-quality videos, can lead people to greatly overestimate how much they have learned by mistaking the *ease of processing* (fluency) for the *depth of learning* (Carpenter et al. 2013).

AI, particularly generative AI, is *fluency on demand*. It produces text (and increasingly multimedia resources) that is coherent, confident, and articulate. This fluency serves as a powerful but misleading metacognitive cue. This issue is precisely what Zhang and Xu (2025) identify as the paradox of self-efficacy: AI use *increased* students' task-related confidence and self-efficacy, but it also *intensified* their technological dependence. The AI-induced feeling of competence, i.e., the illusion of competence, signals to the learner that deep cognitive engagement is no longer necessary. Lee et al. (2025) found evidence to support this theory. They found that greater *confidence* in the capabilities of AI was directly linked to *reduced* critical thinking in the user.

A 2025 qualitative study by Patac and Patac provides the students' perspective, finding that students *perceive* ChatGPT as reducing *all* forms of cognitive load, including the "good" (intrinsic) load. Further evidence from the OECD supports this finding, showing that students in a European sample primarily use AI to make their learning easier and more efficient (OECD 2026). This is not an irrational choice; it is a rational one if the student's goal, driven by a desire for efficiency (Zhai et al. 2024), has shifted from learning to task completion. This situation creates a vicious cycle: a student seeks efficiency; the fluency of AI-generated responses creates an illusion of competence; the illusion triggers metacognitive laziness; this leads to more outsourcing, which erodes the student's actual knowledge base, making them *more* dependent on the tool and *less* able to judge its output in the future.

The pedagogical solution: From cognitive atrophy to augmentation

The evidence for detrimental offloading is compelling, but it is not deterministic. The technology itself does not seal the outcome. As Kalyuga and Plass (2025) argue in their update to Cognitive Load Theory, the goal of the student matters, as does their motivation. Much of the research to date has yet to grapple with this added complexity. As Weidlich et al. (2025) argue in their methodological critique, much of the research on AI is an "effect in search of a cause," because it conflates the *medium* (the AI tool) with the *instructional method* (the pedagogy). They argue that *pedagogy* is the causal factor, and the research provides clear, evidence-informed pathways for a pedagogy that mitigates the risks posed by AI outsourcing. Not only is goal setting (a component of self-regulated learning) critical here, but other aspects of students' capability to manage their own learning and thinking also need to be supported through their learning experiences.

That said, the pedagogical design of the tools used in instruction can influence whether students engage in positive or negative cognitive offloading. AI tools vary in their integration of educational evidence and construction of learning pathways and they increasingly form a mediating layer between curriculum and delivery in the classroom (Loble & Stephens 2024a).

Path 1: Beneficial offloading and load reduction instruction

AI can be used for *beneficial* offloading, managing *extraneous* load to free resources for *intrinsic* learning. This requires an explicit pedagogical framework. Martin et al. (2025) provide such a framework, "Load Reduction Instruction" (LRI), which adapts explicit instruction principles useful for managing learning with AI. Drawing on this model, AI can be used to provide scaffolding, structured practice, and feedback, all aimed at managing the cognitive burden on the learner and enabling progressive independence (Martin et al. 2025), in other words, helping students to become better self-regulated learners. As validated by Hong et al. (2025), students *explicitly taught* this cognitive offload instruction model (offloading low-order writing tasks) showed *significantly greater gains in critical thinking*.

Path 2: Scaffolding metacognition to counter laziness

The more profound solution is one that directly tackles the core problem of metacognitive laziness (Fan et al. 2024). If the *problem* is that the convenience of AI encourages learners to abdicate their metacognitive responsibilities, the *solution* is to design AI interactions that explicitly *demand* and *scaffold* those responsibilities. While these design parameters may not be within the direct control of educators, these kinds of prompts can be used in a wide range of scenarios to help students develop these capabilities. It will be helpful when AI tools increasingly have these capabilities built in, but technology is not required for teachers to use these approaches.

A compelling cluster of recent studies demonstrates the success of this approach, building on a longer history of metacognitive prompting research:

- Xu et al. (2025) found that integrating *metacognitive prompts* into an AI environment significantly enhanced learner capability, particularly in terms of self-regulated learning.
- Singh et al. (2025) found that metacognitive prompts designed to make users pause, reflect, and assess their understanding led to more active engagement, broader topic exploration, and deeper inquiry during AI-based search.
- Li et al. (2025) used a progressive prompting intervention, where the AI provides gradually fading scaffolds, and found that it significantly improved both learning achievement and critical thinking skills.

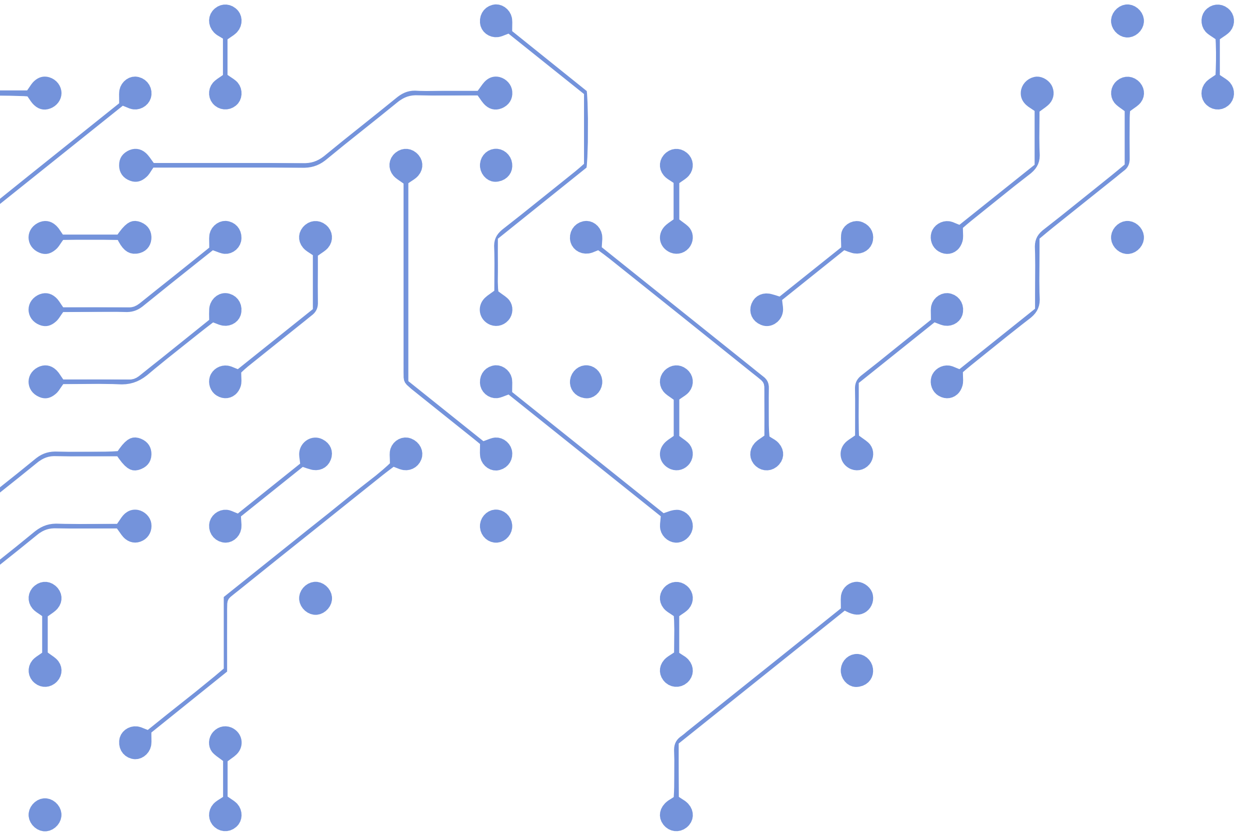
These studies show that the passivity induced by AI can be overcome with explicit metacognitive interventions. Technology is not needed to achieve this. This pedagogy resolves a key conflict: Darvishi et al. (2024) found that simply *adding* SRL prompts failed because the dominance of AI *overrode* them. The successful interventions (e.g., Xu et al. 2025; Singh et al. 2025) were *integrated* and *non-optional*, forcing the metacognitive pause. Many learning activities, with or without technology, can be implemented to achieve this goal.

Path 3: Designing AI as a cognitive mirror and verification partner

The most advanced pedagogical and technological design solutions shift the fundamental role of AI from an answer oracle (which *invites* passive outsourcing) to a tool that *provokes* intrinsic load.

- **AI as cognitive mirror:** Tomisu et al. (2025) propose this framework. The AI is engineered as a teachable novice with a pedagogically useful deficit. It feigns confusion and asks clarifying questions, *forcing* the human learner into the effortful, generative act of explanation and reflection (the "Protégé Effect"), thus *triggering* the generation effect (Duplice 2025).
- **AI as Socratic partner:** Monzon and Hays (2025) propose using AI to *create* desirable difficulties. Instead of *bypassing* effort, the AI is used as a cognitive partner to *generate* retrieval-practice questions, case studies, and Socratic dialogues that *force* the effortful processing required for durable learning.
- **AI as verification partner:** Grace (2025) proposes a model of intelligence equilibrium where the human *maintains primary cognitive agency* and continuously evaluates and corrects the AI output, guided by a verification mindset.

As synthesised by Helal and colleagues (2025), the impact of AI is entirely dependent on the pedagogical design. The default, unstructured answer oracle activates cognitive inhibitors like automation bias and quick-solution dependence. In contrast, intentional, pedagogically sound design activates cognitive mediators like self-regulated learning and metacognitive critique.



> Humans still learn more effectively from and with other humans. By augmenting the teacher, we empower the human expert who is best placed to manage the complex, relational work of co-regulating learning, managing cognitive load, and building the evaluative judgement, self-regulated learning, and metacognition that students need.

Implications for curriculum, practice, and equity

The analysis of cognitive risk points to a significant and novel equity risk. The default, unstructured use of student-facing AI will almost certainly *widen* existing equity gaps, exacerbating existing digital divides (Loble & Hawcroft 2022). However, a *teacher-facing* application of AI offers a robust, scalable, and equitable path forward.

A new metacognitive equity gap

The cognitive risks identified in this report (the performance paradox, metacognitive laziness, and the illusion of competence) are not distributed equally. The research is clear that these negative impacts disproportionately affect *novices*; those who lack the very domain knowledge needed to critically evaluate the output of AI systems (Bastani et al. 2025; Singh et al.,2025). Furthermore, the students most likely to be tempted by the convenience of AI in learning are those who enter the task with weaker self-regulation and metacognitive skills (Darvishi et al. 2024). Gerlich (2025) also found that younger participants showed higher dependence and lower critical thinking.

This suggests a powerful and worrying "Matthew Effect" (Perc 2014) for an AI-mediated world. Students who already possess high levels of domain knowledge and strong metacognitive skills will be able to leverage AI for *beneficial* offloading (as in Hong et al. 2025) and accelerate their learning. Conversely, students *without* these skills, often those already experiencing disadvantage, will be susceptible to *detrimental* offloading, falling prey to the illusion of competence and bypassing the very learning they need. Left unstructured, AI will almost certainly widen existing equity divides (see Loble & Stephens 2024). Explicit instruction on how to manage the risks posed by AI to long-term learning may help but is likely to be insufficient to address this problem at scale. AI 'literacy' is useful but is not intended to be a robust pedagogical response to cognitive outsourcing.

Augmenting the teacher to scale expertise

While the risk of student-facing AI is high, the research suggests a more promising and equitable path. The most powerful use of AI in K–12 education may not be to replace or bypass the teacher with AI tutors, but to *augment* the teachers themselves. Instead of giving the powerful, high-risk tool to the *novice learner*, this model gives it to the *expert teacher* to augment their practice.

Three recent studies demonstrate this potential:

1. **Scaling interventions**: A large, randomised evaluation by Batt et al. (2024) (n=4,000) tested a high-dosage tutoring model that combined an in-person tutor with computer-assisted learning (Khan Academy). This hybrid model produced learning gains (0.23 standard deviations) almost as large as a resource-intensive human-only model, but at a 30% lower cost (Batt et al. 2024). This suggests a viable path for *scaling* high-quality human support with technology.
2. **Scaling expertise**: A 2024 randomised control trial by Wang et al. introduced a "Tutor CoPilot", an AI designed to assist the *tutor*, not the student. This tool provided real-time, expert-like pedagogical guidance to tutors. The study found that this approach significantly improved student pass rates and that the benefits were *greatest* for students of less-experienced or lower-rated tutors. The AI successfully *scaled expertise*, augmenting the human teacher and making them more effective, all at a 165-fold cost reduction compared to traditional professional development (Wang et al. 2024).
3. **Integrating teacher expertise as part of AI design**: A small randomised control trial across five UK high schools tested three tutoring approaches: teacher-controlled chatbot tutoring (chatbot responses could be altered by the teachers); human tutoring; and standardised, pre-written prompts. Students made similar progress whether using the teacher-controlled AI or solely human tutoring, and both groups progressed further than those using generalised hints. Supervising teachers highlighted the teaching-guided chatbot's strength in producing questions that prompted deeper reflection, with results suggesting 'pedagogically fine-tuned AI tutoring systems' offer significant promise (LearnLM Team Google & Eedi 2025).

Further possibilities lie in the emerging use cases for specifically developed AI agents that could function as teachers' aides, of sorts. These agents would likely serve a very different function from human teacher aides, who will continue to play a critical role in education. At the time of writing, the development of AI agents is rapid and unpredictable. Nonetheless, the limited work carried out thus far indicates that this form of teacher augmentation has potential.

These examples represent a foundational alternative form of AI implementation for consideration. The safest, most equitable, and cognitively sound application of AI in schools may be a *teacher-facing* model, especially when partnered with evidence-backed teacher support and with well-designed, trustworthy AI tools. Even with all the research carried out in recent years on AI in education, one fact remains undeniable: humans still learn more effectively from and with other humans. By augmenting the teacher, we empower the human expert who is best placed to manage the complex, relational work of co-regulating learning, managing cognitive load, and building the evaluative judgement, self-regulated learning, and metacognition that students need (Walton et al. 2025).

Conclusion

The challenge of AI is fundamentally pedagogical, not technological (Weidlich et al. 2025). Its effect on learning is determined by the instructional context and the learner's goals.

That is not to say we cannot expect substantial development in AI tools specifically designed for educational contexts. Indeed, many of these tools are emerging (OECD 2026). Nor is it the case that the onus will fall completely on teachers to manage these complex dynamics in their classroom. Nonetheless, the evidence presented in this report highlights the cognitive tension between the zone of overconfidence, driven by fluency, and cognitive ease (Lodge 2023; Stadler et al. 2024), and the zone of maximised learning, which requires effortful, desirable difficulties (de Bruin et al. 2023; Duplice 2025).

The stark warning from Hamilton et al. (2023) of a future of "de-education" and a "mass downgrading of humans" represents the ultimate risk of ubiquitous, frictionless cognitive outsourcing. However, the psychological science and educational technology research detailed in this report provides a clear, pedagogical path to avoid that outcome. That path is not to increase so-called 'AI literacy', which could divert valuable resources from supporting teachers in effectively using AI. Simplified approaches to these issues also belie the system-level changes that will be required as these technologies increasingly impact on how we all live and learn.

The educational imperative is not to *protect* students from a world where cognitive (and metacognitive) offloading is the norm, but to *prepare* them for it. The key components of this preparation are not focused on the technologies themselves but on deep knowledge and adaptive, transferable skills.

This preparation has two non-negotiable components:

1. Arming students with the deep, domain-specific *knowledge* and analytical thinking capabilities they need to think critically *about* the fluent, unreliable output AI can and does generate (Willingham, 2019; Tricot & Sweller, 2014).
2. Fostering the robust *metacognitive judgement* and self-regulated learning skills they need to think critically *with* it, avoiding detrimental offloading and abdication (Xu et al., 2025; Walton et al., 2025).

The pedagogical strategies outlined in this report, from explicit Load Reduction Instruction to AI-driven metacognitive prompts, and the teacher-augmentation models offer an explicit, practical, and hopeful way forward. That is not to say that the suggestions provided here will be straightforward to implement in educational institutions and systems already under significant strain on multiple fronts. Much research and proof-of-concept work is still required, though at least part of the pathway forward has started with the development of educational and human rights standards to guide AI used in Australian classrooms; standards that now should be more widely adopted. What this report provides is a deeply evidence-informed framework for ensuring that AI is harnessed to *augment*, rather than *atrophy*, human learning and critical thought.

References

Australian Curriculum, Assessment and Reporting Authority (2024). *National Report on Schooling in Australia*, accessed 3 February 2026.

Baddeley, A. (1986). *Working memory*. Oxford University Press.

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 of the United States of America, 122*(26), e2422633122. https://doi.org/10.1073/pnas.2422633122

Batt, M. P., Guryan, J., Khan, S. A., LaForest-Tucker, M., & Mishra, B. (2024). *Can technology facilitate scale? Evidence from a randomized evaluation of high dosage tutoring* (NBER Working Paper No. 32510). National Bureau of Economic Research. https://www.nber.org/papers/w32510

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education, 49*(6), 893–905. https://doi.org/10.1080/02602938.2024.2335321

Bjork, E. L., & Bjork, R. A. (2015). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society*. Worth.

Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review, 20*(6), 1350–1356.

Centre for Education Statistics and Evaluation (CESE). (2017). *Cognitive load theory: Research that teachers really need to understand*. NSW Department of Education.

Cowan, J. (2010). Developing the ability for making evaluative judgements. *Teaching in Higher Education, 15*(3), 323–334.

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education, 210*, 104967. https://doi.org/10.1016/j.compedu.2023.104967

David, L., Vassena, E., & Bijleveld, E. (2024). The unpleasantness of thinking: A meta-analytic review of the association between mental effort and negative affect. *Psychological Bulletin, 150*(9), 1070–1093. https://doi.org/10.1037/bul0000443

de Bruin, A. B. H., Biwer, F., Hui, L., Onan, E., David, L., & Wiradhany, W. (2023). Worth the effort: The Start and Stick to Desirable Difficulties (S2D2) framework. *Educational Psychology Review, 35*(2), Article 41. https://doi.org/10.1007/s10648-023-09766-w

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2024). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 223*, 105224. https://doi.org/10.1016/j.compedu.2024.105224

Dron, J. (2022). Educational technology: What it is and how it works. *AI & Society* (37), 155–166. https://doi.org/10.1007/s00146-021-01195-z

Dunn, T. L., & Risko, E. F. (2016). Toward a metacognitive account of cognitive offloading. *Cognitive Science, 40*(5), 1080–1127. https://doi.org/10.1111/cogs.12273

Duplice, J. (2025). Generation and L2 vocabulary learning: A classroom action study on the efficacy of the generation desirable difficulty in learning L2 vocabulary. *Vocabulary Learning and Instruction, 14*(1), Article 102482. https://doi.org/10.29140/vli.v14n1.102482

Ejaz, A., Farhan, M., & Alessi Longa, F. E. (2025). AI and cognitive load: How reliance on AI tools (Chatgpt, Etc.) affects critical thinking. *Research Journal of Psychology, 3*(4), 1–10. https://doi.org/10.59075/rjs.v3i4.245

Ellis, C., & Murdoch, K. (2024). The educational integrity enforcement pyramid: a new framework for challenging and responding to student cheating. *Assessment & Evaluation in Higher Education, 49*(7), 923–934. https://doi.org/10.1080/02602938.2024.2329167

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2024). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Fawns, T. (2022). An entangled pedagogy: Looking beyond the pedagogy–technology dichotomy. *Postdigital Science and Education, 4*(3), 711–728. https://doi.org/10.1007/s42438-022-00302-7

Fogg, B. J. (2003). *Persuasive technology: Using computers to change what we think and do*. Morgan Kaufmann, Amsterdam. https://doi.org/10.1145/764008.763957

Geary, D. C. (1995). Reflections of evolution and culture in children's cognition: Implications for mathematical development and instruction. *American Psychologist*, *50*(1), 24.

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

Grace, S. L. (2025). Finding equilibrium: An integrative approach to balancing human and artificial intelligence in legal research. *Legal Reference Services Quarterly, 44*(1), 1–32. https://doi.org/10.1080/0270319X.2025.2534229

Grinschgl, S., Papenmeier, F., & Meyerhoff, H. S. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology, 74*(9), 1477–1496. https://doi.org/10.1177/17470218211008060

Hamilton, A., Wiliam, D., & Hattie, J. (2023, August 8). *The future of AI in education: 13 things we can do to minimize the damage*. https://www.researchgate.net/publication/373108877

Han, X., Peng, H., & Liu, M. (2025). The impact of GenAI on learning outcomes: A systematic review and meta-analysis of experimental studies. *Educational Research Review, 45*, 100714. https://doi.org/10.1016/j.edurev.2025.100714

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*(7), 325–334. https://doi.org/10.30564/fls.v7i7.10072

Horvath, J. C. (2026). *The digital delusion*: *How classroom technology harms our kids' learning and how to help them thrive again.* LME Global.

Iqbal, J., Hashmi, Z. F., Asghar, M. Z., & Abid, M. N. (2025). Generative AI tool use enhances academic achievement in sustainable education through shared metacognition and cognitive offloading among preservice teachers. *Scientific Reports, 15*(1), 16610. https://doi.org/10.1038/s41598-025-01676-x

Jensen B, Ross M, Collett M, Murnane M, and Pearson E (2023) Fixing the hole in Australian education: The Australian curriculum benchmarked against the best. Learning First, accessed 3 February 2026.

Kalyuga, S., & Plass, J. L. (2025). *Rethinking cognitive load theory.* Oxford University Press.

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports, 15*(1), Article 17458. https://doi.org/10.1038/s41598-025-97652-6

Kolhatin, A. O. (2025). From automation to augmentation: A human-centered framework for generative AI in adaptive educational content creation. In *Proceedings of the 8th International Workshop on Augmented Reality in Education (AREdu 2025)* (CEUR Workshop Proceedings, Vol. 4060, pp. 58–69). CEUR-WS.org.

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes,P. (2025). *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task* [Preprint]. arXiv. https://arxiv.org/abs/2506.08872

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Lademann, J., Henze, J., & Becker-Genschow, S. (2025). Augmenting learning environments using AI custom chatbots: Effects on learning performance, cognitive load, and affective variables. *Physical Review Physics Education Research, 21*(1), 010147. https://doi.org/10.1103/PhysRevPhysEducRes.21.010147

Lee, H. P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI conference on human factors in computing systems* (pp. 1-22). ACM. https://doi.org/10.1145/3706598.3713778

Leibovitch, Y. M., Brown, D. J., Pearson, S., Mazzola, C., Ellerton, P. J., & Watt, G. (2024). Investigating the impact of critical thinking instruction on writing performance: A multilevel modelling analysis of relative gain data in the Australian national assessment program. *Thinking Skills and Creativity*, *53*, 101546. https://doi.org/10.1016/j.tsc.2024.101546

Li, C.-J., Huang, G.-J., Chang, C.-Y., & Su, H.-C. (2025). Generative AI supported progressive prompting for professional training: Effects on learning achievement, critical thinking, and cognitive load. *British Journal of Educational Technology, 56*(6), 2550–2572. https://doi.org/10.1111/bjet.13594

Loble, L., & Hawcroft, A. (2022). *Shaping AI and edtech to tackle Australia's learning divide*. The University of Technology Sydney. https://doi.org/10.57956/kxye-qd93

Loble, L., & Stephens, K. (2024). *Securing digital equity in Australian education*. The University of Technology Sydney. https://doi.org/10.57956/rrpc-5708

Loble, L., & Stephens, K. (2024a). *Towards high quality in Australian educational technology.* The University of Technology Sydney. https://doi.org/10.57956/8dbd-yj25

Lodge, J. M. (2023). Misjudgements of learning in digital environments. In S. Y. L. Chye & B. L. Chua (Eds.), *Pedagogy and psychology in digital education* (pp. 239–247). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-2107-2_14

Lodge, J. M., Howard, S., Bearman, M., Dawson, P., & Associates (2023a). *Assessment reform for the age of Artificial Intelligence*. Tertiary Education Quality and Standards Agency.

Lodge, J. M., Kennedy, G. & Hattie, J. A. C. (2018a). Understanding, assessing and enhancing student evaluative judgement in digital environments. In D. Boud, R. Ajjawi, P. Dawson & J. Tai (Eds.) *Developing evaluative judgement in higher education: Assessment for knowing and producing quality work.* Routledge.

Lodge, J. M., Kennedy, G. E., Lockyer, L., Arguel, A. & Pachman, M. (2018). Understanding difficulties and resulting confusion in learning: An integrative review. *Frontiers in Education, 3* (49). https://doi.org/10.3389/feduc.2018.00049

Lodge, J. M., Yang, S., Furze, L., & Dawson, P. (2023). It's not like a calculator, so what is the relationship between learners and generative artificial intelligence? *Learning: Research and Practice, 9*(2), 117–124. https://doi.org/10.1080/23735082.2023.2261106

Martin, A. J., Collie, R. J., Kennett, R., Liu, D., Ginns, P., Sudimantara, L. B., Dewi, E. W., & Rüschenpöhler, L. G. (2025). Integrating generative AI and load reduction instruction to individualise and optimise students' learning. *Learning and Individual Differences, 121*, Article 102723. https://doi.org/10.1016/j.lindif.2025.102723

Monzon, N., & Hays, F. A. (2025). Leveraging generative artificial intelligence to improve motivation and retrieval in higher education learners. *JMIR Medical Education, 11*, e59210. https://doi.org/10.2196/59210

Helal, M. Y. I., Elgendy, I. A., Albashrawi, M. A., Dwivedi, Y. K., Al-Ahmadi, M. S., & Jeon, I. (2025). The impact of generative AI on critical thinking skills: A systematic review, conceptual framework and future research directions. *Information Discovery and Delivery*. Advance online publication. https://doi.org/10.1108/IDD-05-2025-0125

LearnLM team Google and Eedi (2025). AI tutoring can safely and effectively support students: An exploratory RCT in UK classrooms. ArXiv:2512.23633 [cs.CY]. https://doi.org/10.48550/arXiv.2512.23633

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition In J. Metcalfe, & A. P. Shimamura (Eds.). *Metacognition: Knowing about knowing*. MIT Press.

NSW (2024). *Expansion of AI education tool trial delivers more time for teachers in our classrooms*. NSW Department of Education, accessed 3 February 2026.

Oakley, B., Johnston, M., Chen, K.Z., Jung, E., & Sejnowski, T.J. (2025). The Memory Paradox: Why Our Brains Need Knowledge in an Age of AI. ArXiv:2506.11015v2 [cs.CY]. https://doi.org/10.48550/arXiv.2506.11015

OECD (2025), *Results from TALIS 2024: The State of Teaching*. TALIS OECD Publishing. https://doi.org/10.1787/90df6235-en

OECD. (2026). *OECD digital education outlook 2026: Exploring effective uses of generative AI in education*. OECD Publishing. https://doi.org/10.1787/062a7394-en

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

Patac, L. P., & Patac, A. V., Jr. (2025). Using ChatGPT for academic support: Managing cognitive load and enhancing learning efficiency – A phenomenological approach. *Social Sciences & Humanities Open, 11*, Article 101301. https://doi.org/10.1016/j.ssaho.2025.101301

Perc, M. (2014). The Matthew effect in empirical data. *Journal of the Royal Society Interface*, *11*(98), 20140378. https://doi.org/10.1098/rsif.2014.0378

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

Sample, A. (2020). Historical development of definitions of information literacy: A literature review of selected resources. *The Journal of Academic Librarianship*, *46*(2), 102116. https://doi.org/10.1016/j.acalib.2020.102116

Singh, A., Taneja, K., Guan, Z., & Ghosh, A. (2025). Protecting human cognition in the age of AI. arXiv preprint arXiv:2502.12447. https://arxiv.org/abs/2502.12447

Soderstrom N. C. & Bjork R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science, 10* (2), 176–199. https://doi.org/10.1177/1745691615569000

Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior, 160*, 108386. https://doi.org/10.1016/j.chb.2024.108386

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*, 261–292. https://doi.org/10.1007/s10648-019-09465-5

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3

Thomas, J., McCosker, A., Parkinson, S., Hegarty, K., Featherstone, D., Kennedy, J., Ormond-Parker, L. Morrison, K., Rea, H., & Ganley, L. *Measuring Australia's Digital Divide: 2025 Australian Digital Inclusion Index*. Melbourne: ARC Centre of Excellence for Automated Decision-Making and Society, RMIT University, Swinburne University of Technology, and Telstra. https://doi.org/10.60836/mtsq-at22

Tomisu, H., Ueda, J., & Yamanaka, T. (2025). The cognitive mirror: A framework for AI-powered metacognition and self-regulated learning. *Frontiers in Education, 10*, Article 1697554. https://doi.org/10.3389/feduc.2025.1697554

Tricot, A., & Sweller, J. (2014). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review, 26*(2), 265–283. https://doi.org/10.1007/s10648-013-9243-1

Trumbo, M., McDaniel, M. A., Hodge, G. K., Jones, A. P., Matzen, L. E., Kittinger, L. I., ... & Clark, V. P. (2021). Is the testing effect ready to be put to work? Evidence from the laboratory to the classroom. *Translational Issues in Psychological Science*, *7*(3), 332.

UNESCO (United Nations Educational, Scientific and Cultural Organization) (2023). *Guidance for Generative AI in Education and Research*. UNESCO, accessed 9 March 2026.

Walton, J., Bearman, M., Crawford, N., Tai, J., & Boud, D. (2025). How university students work on assessment tasks with generative artificial intelligence: Matters of judgement. *Assessment & Evaluation in Higher Education, 50*(5), 705–721. https://doi.org/10.1080/02602938.2025.2570328

Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. *Humanities and Social Sciences Communications, 12*(1), Article 621. https://doi.org/10.1057/s41599-025-04787-y

Wang, R. E., Ribeiro, A. T., Robinson, C. D., Loeb, S., & Demszky, D. (2024). *Tutor CoPilot: A human-AI approach for scaling real-time expertise*. arXiv preprint arXiv:2410.03017. https://arxiv.org/abs/2410.03017

Wang, T., & Lajoie, S. P. (2023). How does cognitive load interact with self-regulated learning? A dynamic and integrative model. *Educational Psychology Review, 35*(3), Article 69. https://doi.org/10.1007/s10648-023-09794-6

Weidlich, J., Gašević, D., Drachsler, H., & Kirschner, P. (2CH25). ChatGPT in education: An effect in search of a cause. *Journal of Computer Assisted Learning, 41*(4), e70105. https://doi.org/10.1111/jcal.70105

Willingham, D. T. (2019). *How to teach critical thinking.* NSW Department of Education. https://education.nsw.gov.au/teaching-and-learning/education-for-a-changing-world/thinking-skills/how-to-teach-critical-thinking

Xu, X., Qiao, L., Cheng, N., Liu, H., & Zhao, W. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *British Journal of Educational Technology, 56*(5), 1842–1863. https://doi.org/10.1111/bjet.13599

Yan, L., Greiff, S., Lodge, J. M., & Gašević, D. (2025). Distinguishing performance gains from learning when using generative AI. *Nature Reviews Psychology, 4*(7), 435–436. https://doi.org/10.1038/s44159-025-00467-5

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour, 8*(10), 1839–1850. https://doi.org/10.1038/s41562-024-02004-5

Yang, K., Raković, M., Liang, Z., Yan, L., Zeng, Z., Fan, Y., Gašević, D., & Chen, G. (2025, March 3–7). Modifying AI, enhancing essays: How active engagement with generative AI boosts writing quality. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference (LAK '25)* (pp. 568–578). ACM. https://doi.org/10.1145/3706468.3706544

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*(1), Article 28. https://doi.org/10.1186/s40561-024-00316-7

Zhang, L., & Xu, J. (2025). The paradox of self-efficacy and technological dependence: Unraveling generative AI's impact on university students' task completion. *The Internet and Higher Education, 65*, 100978. https://doi.org/10.1016/j.iheduc.2024.100978

Appendix: Annotated bibliography

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 of the United States of America, 122*(26), e2422633122. https://doi.org/10.1073/pnas.2422633122

In the study reported in this paper, "nearly 1,000" year 9 to 11 students were given access to generative AI to work through mathematics problems. The AI appeared to facilitate students' problem-solving during learning. However, their learning suffered once the AI was taken away. This suggests that something beyond the problem-solving exercise was being lost in the process. This paper provides critical empirical evidence for the "performance paradox," demonstrating that scaffolded performance with AI does not equate to durable learning.

Batt, M. P., Guryan, J., Khan, S. A., LaForest-Tucker, M., & Mishra, B. (2024). *Can technology facilitate scale? Evidence from a randomized evaluation of high dosage tutoring* (NBER Working Paper No. 32510). National Bureau of Economic Research. https://www.nber.org/papers/w32510

This RCT with 4,000 Chicago and New York City students tested whether computer-assisted learning could reduce high-dosage tutoring costs while maintaining effectiveness. Students alternated between human tutors and Khan Academy in groups of four, reducing per-pupil expenses by 30%. The intervention achieved 0.23 standard deviation gains in math scores, comparable to more expensive pure-tutoring models. The study demonstrates that strategic technology integration (rather than wholesale substitution) can maintain educational effectiveness while addressing cost barriers to scaling evidence-based interventions. The authors claim that the evidence offers a counterpoint to fears about technology replacing teachers.

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education, 210*, 104967. https://doi.org/10.1016/j.compedu.2023.104967

This empirical study investigated the impact of AI assistance on student agency, operationalised as self-regulated learning (SRL) strategies in a peer-feedback context. The results demonstrate a tendency for students to *rely on* rather than *learn from* the AI. When AI assistance was removed, students who employed SRL strategies could not match the performance achieved with AI. Critically, a hybrid human-AI approach, which combined AI assistance with SRL strategy prompts, was *not* more effective than AI assistance alone. The authors conclude that the dominance of the AI support, potentially due to its reduction of cognitive load, may override or supplant the learner's own regulatory processes, posing a significant challenge to fostering student agency.

de Bruin, A. B. H., Biwer, F., Hui, L., Onan, E., David, L., & Wiradhany, W. (2023). Worth the effort: The Start and Stick to Desirable Difficulties (S2D2) framework. *Educational Psychology Review, 35*(2), Article 41. https://doi.org/10.1007/s10648-023-09766-w

This theoretical paper presents a framework for understanding and enhancing the regulation of effort in learning. It is an important skill for ensuring that students both invest the effort required to learn effectively and understand the value of the effort relative to the use of AI to shortcut it. The S2D2 framework provides the theoretical basis for desirable difficulties, which generative AI may encourage students to bypass, to the detriment of their long-term learning.

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2024). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 223*, 105224. https://doi.org/10.1016/j.compedu.2024.105224

A widely reported systematic review and meta-analysis of the impact of generative AI on learning was published in 2024. The review and meta-analysis both indicated that generative AI could have a positive influence on learning and on the reduction of mental effort. This study has generated a lot of debate. In particular, the framing of the results runs counter to the work on desirable difficulties and the importance of intrinsic cognitive load. This review is a key example of the "fast science" critiqued by Weidlich et al. (2025), as its positive findings are likely based on studies that measure short-term performance rather than durable learning.

Dunn, T. L., & Risko, E. F. (2016). Toward a metacognitive account of cognitive offloading. *Cognitive Science, 40*(5), 1080–1127. https://doi.org/10.1111/cogs.12273

This older article includes experimental studies (with small sample sizes) showing that metacognition plays a critical role in determining *when* and *how* to offload cognitive work to an external source. While difficult to interpret due to the lack of power, the study points to the important role metacognition plays in offloading. It provides a foundational pre-AI framework, arguing that offloading is an active, metacognitively guided *choice* based on a cognitive cost-benefit analysis.

Duplice, J. (2025). Generation and L2 vocabulary learning: A classroom action study on the efficacy of the generation desirable difficulty in learning L2 vocabulary. *Vocabulary Learning and Instruction, 14*(1), Article 102482. https://doi.org/10.29140/vli.v14n1.102482

This classroom action study provides a crucial, non-AI baseline for the "desirable difficulties" framework. It investigates the generation effect, a learning method where students must *generate* an answer (e.g., a word from a cue) rather than just passively review it. The study found that this "generation desirable difficulty" led to significantly better long-term vocabulary retention in L2 learners compared to passive study. This paper is significant as it provides empirical evidence for the *specific cognitive process* that ubiquitous AI tools, by providing answers directly, encourage learners to bypass, thereby offering a direct mechanism to explain the performance paradox.

Ejaz, A., Farhan, M., & Alessi Longa, F. E. (2025). AI and cognitive load: How reliance on AI tools (Chatgpt, Etc.) affects critical thinking. *Research Journal of Psychology, 3*(4), 1–10. https://doi.org/10.59075/rjs.v3i4.245

This quantitative survey study (n=350 university students) investigated the relationship between AI tool reliance, cognitive load, and critical thinking. The findings reveal a significant negative correlation between frequent AI use and self-assessed critical thinking skills. This relationship was mediated by cognitive load: students who used AI more frequently reported a *lower* cognitive load during complex tasks, which in turn was associated with *lower* critical thinking. The study also found a strong *positive* correlation between higher cognitive load and better critical thinking, empirically supporting the desirable difficulties framework. The authors posit this is direct evidence of detrimental cognitive offloading, where AI's convenience discourages deep cognitive engagement.

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2024). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56*(2), 489–530. https://doi.org/10.1111/bjet.13544

This paper introduces the critical concept of "metacognitive laziness". In a randomised experimental study on a writing task, the authors compared learners with support from AI, human experts, and intelligent learning systems. The findings suggest that the convenience of AI can undermine learners' engagement in essential self-regulatory processes, such as planning, monitoring, and revision. This laziness represents a form of detrimental cognitive offloading where the learner abdicates their metacognitive responsibilities to the tool. This paper provides a crucial counterpoint to optimistic portrayals of AI as an SRL scaffold, suggesting that without intentional design (like that in Xu et al., 2025), the default effect of AI is to *supplant*, not support, metacognition.

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

This study provides strong empirical evidence for the harm thesis of the performance paradox. Using a mixed-methods approach with 666 participants, the study found a *significant negative correlation* between the frequency of AI tool usage and critical thinking abilities. Critically, it used mediation analysis to demonstrate that this negative relationship is explained by *increased cognitive offloading*. The study also noted that younger participants showed higher dependence and lower critical thinking, suggesting a vulnerability in developing minds. This paper directly supports the mechanism proposed in Ejaz et al. (2025) and provides a robust, large-sample dataset that counters the "fast science" of performance-only meta-analyses (e.g. Deng et al., 2024).

Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., Järvelä, S., Mavrikis, M., & Rienties, B. (2024). The promise and challenges of generative AI in education. *Behaviour & Information Technology, 44*(11), 2518–2544. https://doi.org/10.1080/0144929X.2024.2394886

This high-level commentary article by leading scholars at the intersection of the learning sciences and AI in education provides a snapshot of the potential and pitfalls of AI use across educational contexts. There is nothing surprising in the paper; the authors argue that substantial further research is needed to develop a thorough understanding of the role that AI can and should play in learning into the future. This cautious consensus serves as a counterpoint to the hype surrounding positive meta-analyses.

Gkintoni, E., Antonopoulou, H., Sortwell, A., & Halkiopoulos, C. (2025). Challenging Cognitive Load Theory: The role of educational neuroscience and artificial intelligence in redefining learning efficacy. *Brain Sciences, 15*(2), Article 203. https://doi.org/10.3390/brainsci15020203

This systematic review connects Cognitive Load Theory, AI, and Educational Neuroscience (EdNeuro). Its primary contribution is to advocate for the use of neurophysiological tools (e.g., EEG, fNIRS) to *objectively measure* cognitive load, moving beyond the self-report scales used in studies like Ejaz et al. (2025). The authors argue that AI-driven adaptive systems, informed by real-time neuroscientific insights into a learner's cognitive state, can optimise instruction. This paper provides a neuroscientific and methodological foundation for the beneficial offloading hypothesis, where AI could (in theory) manage extraneous load with high precision, freeing resources for germane load.

Glickman, M., & Sharot, T. (2025). How human–AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour, 9*(2), 345–359. https://doi.org/10.1038/s41562-024-02077-2

In a set of large studies with over 1,400 participants, the researchers found that ongoing interactions and feedback loops between learners and AI tend to generate a drift towards amplifying biases (usually negative). This is interesting but needs further exploration. In a CLT context, this bias amplification could be seen as introducing a novel form of *extraneous cognitive load* that actively hinders, rather than helps, the learning process.

Grace, S. L. (2025). Finding equilibrium: An integrative approach to balancing human and artificial intelligence in legal research. *Legal Reference Services Quarterly, 44*(1), 1–32. https://doi.org/10.1080/0270319X.2025.2534229

This theoretical article applies cognitive psychology frameworks, including Cognitive Load Theory, distributed cognition, and metacognition, to the use of generative AI in legal research. It warns that while AI promises efficiency, it risks cognitive overload, overreliance, and the erosion of metacognitive habits. The paper introduces the concept of "intelligence equilibrium," a state where AI *complements* rather than *replaces* human analysis. It argues that achieving this equilibrium requires instructional strategies focused on a "verification mindset," where the human (the learner) maintains primary cognitive agency and continuously evaluates and corrects the AI output.

Grinschgl, S., Papenmeier, F., & Meyerhoff, H. S. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology, 74*(9), 1477–1496. https://doi.org/10.1177/17470218211008060

This study, carried out prior to the release of ChatGPT, includes three experiments on cognitive offloading, each with 172 participants. Across these studies, it is evident that the negative effects of offloading are related to the goal orientation of the learner. This pre-AI finding anticipates the goal-driven reframing of Cognitive Load Theory proposed by Kalyuga & Plass (2025), suggesting that *why* a learner offloads is the determinative factor.

Grinschgl, S., & Neubauer, A. C. (2022). Supporting cognition with modern technology: Distributed cognition today and in an AI-enhanced future. *Frontiers in Artificial Intelligence, 5*, Article 908261. https://doi.org/10.3389/frai.2022.908261

This review article covers the potential benefits and risks of using advanced technologies, particularly AI, to offload cognitive tasks. The authors argue that there is a place for offloading to AI, but over-reliance on technologies for offloading remains a significant potential risk. The paper was published in 2022 but seems to accurately predict the performance paradox (boosting performance but diminishing memory) that would be widely reported in 2024–2025.

Hamilton, A., Wiliam, D., & Hattie, J. (2023, August 8). The future of AI in education: 13 things we can do to minimize the damage [Working paper]. https://www.researchgate.net/publication/373108877

This 39-page synthesis by three prominent education scholars examines existential questions about AI's impact on cognitive development and the purpose of education. The authors argue humanity may be at peak humanity, with AI posing fundamental risks to cognitive development and agency. They present four future scenarios ranging from AI curtailment to transhumanism, each with profound educational implications. The paper's 13 regulatory recommendations include assuming AGI arrival within two years, establishing global frameworks, and restricting student access pending risk assessment. This work stands out for addressing existential concerns beyond plagiarism, providing actionable policy guidance at a critical juncture in AI development.

Han, X., Peng, H., & Liu, M. (2025). The impact of GenAI on learning outcomes: A systematic review and meta-analysis of experimental studies. *Educational Research Review, 45*, 100714. https://doi.org/10.1016/j.edurev.2025.100714

Another systematic review and meta-analysis, this time from 2025. The findings from this one also support the role of generative AI in improving learning. The authors also found significant moderators, particularly the age of the cohort, with effects being larger in primary and secondary school contexts. The findings need to be treated with caution, as the studies included in the analysis are often underpowered and use only testing (i.e., short-term performance) to determine the effect.

Han, Z., Ying, R., Huang, C., Tsai, C. C., Wang, X., & He, T. (2025). Identifying students' metacognition patterns by their needs for cognitive closure in human-GenAI collaboration. *Computers & Education, 225*, 105422. https://doi.org/10.1016/j.compedu.2025.105422

This study, with 82 undergraduates split into small groups, found a relationship between the "Need for Cognitive Closure" (a desire for unambiguous answers) and AI use patterns. The results suggest a complex interaction between generative AI use and motivational factors that requires further research. This work provides a potential psychological mechanism for detrimental cognitive offloading: a high need for closure leads to passive acceptance of AI answers.

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*(7), 325–334. https://doi.org/10.30564/fls.v7i7.10072

This paper provides a direct, pedagogical intervention to distinguish beneficial from detrimental offloading. This quasi-experimental study (n=240) tested a cognitive offload instruction model where students were *explicitly taught* to delegate lower-order writing tasks (e.g., brainstorming, grammar) to AI, compelling them to focus on higher-order analysis, evaluation, and reflection. The results showed that this experimental group had *significantly greater improvements* in critical thinking assessments and essay quality compared to controls. This study is a critical synthesis, empirically demonstrating the beneficial offloading hypothesis (Iqbal et al., 2025) and confirming the central thesis of Weidlich et al. (2025) that *pedagogy*, not the medium, drives learning.

Iqbal, J., Hashmi, Z. F., Asghar, M. Z., & Abid, M. N. (2025). Generative AI tool use enhances academic achievement in sustainable education through shared metacognition and cognitive offloading among preservice teachers. *Scientific Reports, 15*(1), Article 16610. https://doi.org/10.1038/s41598-025-01676-x

This quantitative study (n=465) used a time-lag survey design to model the factors influencing academic achievement among preservice teachers using AI. The study found that AI tool use was positively associated with achievement. Critically, this relationship was *fully mediated* by two key cognitive factors: *shared metacognition (SMC)* and *cognitive offloading (COL)*. Both SMC and COL were found to be strong positive mediators, indicating that AI tools enhance learning *when* they facilitate collaborative metacognition and effectively offload lower-order cognitive tasks (i.e., extraneous load), thereby freeing resources for higher-order learning.

Jose, B., Cherian, J., Verghis, A. M., Varghise, S. M., S, M., & Joseph, S. (2025). The cognitive paradox of AI in education: Between enhancement and erosion. *Frontiers in Psychology, 16*, Article 1550621. https://doi.org/10.3389/fpsyg.2025.1550621

This theoretical review paper explicitly explores the cognitive paradox of AI in education, arguing that AI can simultaneously enhance lower-order skills (like recall) while eroding higher-order critical thinking. The authors use Cognitive Load Theory and Self-Determination Theory to frame this conflict. They posit that while AI can reduce cognitive load, an over-reliance on it as a substitute for mental effort (i.e., cognitive offloading) can compromise learner autonomy and stifle the development of analytical skills. The paper calls for pedagogically informed AI integration that balances technological efficiency with the cognitive effort essential for deep, meaningful learning.

Kalyuga, S., & Plass, J. L. (2025). *Rethinking cognitive load theory*. Oxford University Press.

This foundational *book* proposes a major reframing of Cognitive Load Theory (CLT), arguing for a more learner-centred framework. The authors make two major modifications: (1) They return to a two-load model (intrinsic, extraneous) and discard germane cognitive load as a separate construct, arguing it is conceptually problematic. (2) They propose an integrated, goal-driven approach where learner activities, motivations, and affective goals—not just schema acquisition—are the focus of analysis. This reframing is essential for AI research, as it suggests the *learner's goal* (e.g., "get the answer" vs. "learn the concept") is the key factor determining whether offloading is beneficial or harmful.

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports, 15*(1), Article 17458. https://doi.org/10.1038/s41598-025-97652-6

This large, randomised control trial tested an AI tutor compared to a set of active learning activities in a cross-sectional design. Participants performed better in a post-test when engaged with the AI tutor and reported higher levels of motivation. The results only apply to knowledge acquisition but show some potential for AI tutors nonetheless. This study is a key example of the performance paradox, as the post-test measures scaffolded performance, not the durable learning that Bastani et al. (2025) found was *harmed*.

Kolhatin, A. O. (2025). From automation to augmentation: A human-centered framework for generative AI in adaptive educational content creation. In *Proceedings of the 8th International Workshop on Augmented Reality in Education (AREdu 2025)* (CEUR Workshop Proceedings, Vol. 4060, pp. 58–69). CEUR-WS.org.

This conference paper reviews implementations of generative AI in education, proposing a human-centred framework that prioritises augmentation over automation. The analysis of empirical evidence identifies common characteristics of successful implementations (25-60% learning improvements), such as pedagogical primacy and human-in-the-loop architectures. Critically, the paper also identifies significant challenges from the literature, including a finding that *cognitive offloading effects reduced independent problem-solving by 35%*. The paper argues for a framework that addresses these risks by maintaining educator oversight and ensuring learning science, not technology, drives deployment.

Kulal, A. (2025). Cognitive risks of AI: Literacy, trust, and critical thinking. *Journal of Computer Information Systems*, 1–13. https://doi.org/10.1080/08874417.2025.2582050

A survey study of 625 postgraduate students in a higher education context in India. Using a moderation analysis, the author found that there was a relationship between AI use and critical thinking. The findings suggest that the more AI is used, the less critical thinking is engaged. The situation is seemingly compounded when participants had high levels of trust in AI systems. Although the author alludes to causal factors, the design of the study is correlational and can therefore only support the result that these factors are related among the students in this relatively niche and homogenous sample.

Lademann, J., Henze, J., & Becker-Genschow, S. (2025). Augmenting learning environments using AI custom chatbots: Effects on learning performance, cognitive load, and affective variables. *Physical Review Physics Education Research, 21*(1), 010147. https://doi.org/10.1103/PhysRevPhysEducRes.21.010147

This study implemented an AI chatbot into a year six class on physics and mathematics in Germany. In a cohort of 214 students, they found that the use of a chatbot enhanced learning and engagement while also lowering cognitive load. The authors took this as evidence that AI chatbots are a useful addition to education, particularly in STEM subjects. It is worth noting here that the primary outcome was an exam. This finding highlights the ambiguity in the field: the tool reduced load and improved *performance*, but it is unknown if it improved *durable learning*.

Lee, H. P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI conference on human factors in computing systems* (pp. 1-22). ACM. https://doi.org/10.1145/3706598.3713778

This survey of knowledge workers (extending the research beyond students) provides a crucial psychological insight into the offloading phenomenon. The study found that greater *confidence* in AI's capabilities was directly linked to *reduced critical thinking*. This suggests a mechanism for the metacognitive laziness described by Fan et al. (2024): users who trust the output of AI are less likely to engage in the effortful cognitive work of verification and analysis. This complements the paradox of self-efficacy (Zhang & Xu, 2025), where the fluency of AI outputs creates an illusion of competence that *dampens* the user's own metacognitive monitoring, leading to detrimental offloading.

Martin, A. J., Collie, R. J., Kennett, R., Liu, D., Ginns, P., Sudimantara, L. B., Dewi, E. W., & Rüschenpöhler, L. G. (2025). Integrating generative AI and load reduction instruction to individualise and optimise students' learning. *Learning and Individual Differences, 121*, Article 102723. https://doi.org/10.1016/j.lindif.2025.102723

This conceptual paper integrates Cognitive Load Theory with Load Reduction Instruction to provide systematic guidance for educational AI implementation. The framework proposes five principles: managing cognitive load types, balancing instruction with discovery, providing appropriate scaffolding, enabling progressive independence, and accommodating the novice-expert continuum. The authors argue that without systematic cognitive load management, poorly designed AI implementations impose problematic burden hampering development. This work bridges established CLT research with emerging AI applications, providing concrete guidance for educators and designers beyond current ad hoc approaches.

Monzon, N., & Hays, F. A. (2025). Leveraging generative artificial intelligence to improve motivation and retrieval in higher education learners. *JMIR Medical Education, 11*, e59210. https://doi.org/10.2196/59210

This viewpoint article from the medical education field argues for a constructive reframing of AI around the desirable difficulties framework. The authors posit that instead of just being a tool that *bypasses* effort (Bastani et al., 2025), AI can be used as a cognitive partner to *create* desirable difficulties. They propose using AI to generate retrieval-practice questions, case studies, and Socratic dialogues that *force* the effortful processing required for durable learning. This provides a clear, positive pedagogical model that directly opposes the performance paradox and aligns with the S2D2 framework (de Bruin et al., 2023).

Oakley, B., Johnston, M., Chen, K.Z., Jung, E., & Sejnowski, T.J. (2025). The Memory Paradox: Why Our Brains Need Knowledge in an Age of AI. ArXiv:2506.11015v2 [cs.CY]. https://doi.org/10.48550/arXiv.2506.11015

This chapter provides comprehensive and accessible insight to the neuroscience, learning theory and memory systems research that explains why explicit knowledge acquisition, practice and memory remain essential and strengthen learning, especially in the context of growing AI use. Surveying research, the authors find that technological dependency, cognitive offloading and education models that de-emphasise procedure and memory can impair consolidation of the mental structures (schemata) necessary for critical thinking, expertise and intuitive mastery, and may provide an explanation for population-level trends in cognitive decline. They outline an educational approach that balances utilisation of AI technologies with a pedagogical focus on building strong knowledge and metacognitive foundations.

Pan, Y., Liu, Y., Zhang, Z., Liu, S., Song, Y., Wang, S., Wu, Y., Zhang, Y., Wang, Z., & Yu, Y. (2025). LLM-powered multi-agent framework for goal-oriented learning in intelligent tutoring system. In *Proceedings of the ACM Web Conference 2025 (WWW '25)*. ACM. https://doi.org/10.48550/arXiv.2501.15749

This paper presents a technical and architectural solution aligned with the goal-driven reframing of CLT (Kalyuga & Plass, 2025). The authors developed "GenMentor," an intelligent tutoring system (ITS) framework that is explicitly goal oriented. The system uses multiple LLM agents to identify a learner's *skill gaps* in relation to their *specific goals*, and then provides personalised, adaptive learning paths. This paper is significant because it represents the engineering application of the goal-driven theory, contrasting sharply with the use of generic, non-goal-oriented tools (like the one used in Bastani et al., 2025), which leads to detrimental offloading.

Patac, L. P., & Patac, A. V., Jr. (2025). Using ChatGPT for academic support: Managing cognitive load and enhancing learning efficiency – A phenomenological approach. *Social Sciences & Humanities Open, 11*, Article 101301. https://doi.org/10.1016/j.ssaho.2025.101301

This qualitative study provides the *learner's perspective* on the cognitive load mechanism. Through a phenomenological approach, the authors found that students *perceive* ChatGPT as reducing all forms of cognitive load. This finding is critical: it explains *why* students engage in detrimental offloading (Zhai et al., 2024). They are making a rational choice based on their *perception* of reduced effort (a feeling of fluency), which aligns with the goal-driven framework (Kalyuga & Plass, 2025) where the goal of task efficiency overrides the (more effortful) goal of learning.

Singh, A., Guan, Z., & Rieh, S. Y. (2025). Enhancing critical thinking in generative AI search with metacognitive prompts. *Proceedings of the Association for Information Science and Technology, 62*(1), 672–684. https://doi.org/10.1002/pra2.1287

This small-scale user study (n=40) provides another direct, empirical test of the metacognitive support solution. The study examined the impact of metacognitive prompts on critical thinking during AI-based search. The prompts were designed to make users pause, reflect, and assess their understanding. The results showed that these prompts led to *more active engagement*, *broader topic exploration*, and *deeper inquiry*. This study serves as a direct, practical solution to the problems of passive acceptance and metacognitive laziness (Fan et al., 2024), demonstrating (like Xu et al., 2025) that AI-driven passivity can be overcome with *explicit* metacognitive interventions.

Singh, A., Taneja, K., Guan, Z., & Ghosh, A. (2025; preprint). Protecting human cognition in the age of AI. arXiv preprint arXiv:2502.12447. https://arxiv.org/abs/2502.12447

This comprehensive synthesis applies Bloom's Taxonomy and Dewey's reflective thought framework to analyse cognitive impacts of AI on novice learners. The review reveals concerning patterns: AI shifts users from active information seeking to passive consumption, causes over-reliance even when contradicting personal reasoning, and improves short-term performance while failing to boost knowledge transfer. Students consistently demonstrate metacognitive laziness, mistaking task ease for understanding. The paper argues that AI bypasses crucial cognitive processes necessary for deep learning, providing mechanistic explanations through classical educational psychology theories. Recommendations address educators, test designers, and tool developers on how to minimise cognitive offloading while preserving productive struggle.

Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior, 160*, 108386. https://doi.org/10.1016/j.chb.2024.108386

An important study demonstrating that the effect of generative AI use on cognitive load and performance is complicated. Ninety-one university students completed an assigned task using an LLM-based technology. Doing so led to reductions in cognitive load but *no tangible differences in performance*, aligning with the idea of a performance paradox. The cognitive ease came at the cost of compromising depth. This paper effectively severs the assumed link that a reduction in cognitive load automatically leads to better learning.

Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A. E., Sarkar, A., Sellen, A., & Rintel, S. (2024). The metacognitive demands and opportunities of generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1–24). ACM. https://doi.org/10.1145/3613904.3642902

This non-empirical review article argues that generative AI systems used in learning must incorporate metacognitive support. Their argument is built on the idea that our metacognitive systems are the key to maximising human-machine relationships for learning and for work. The authors reframe the problem, suggesting AI creates *new* metacognitive demands: learners must now monitor not only their own cognition but also the AI's output and the interaction between the two.

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

A survey study was carried out on 580 students in a Chinese higher education context. The study focused on four factors: AI use/dependence, critical thinking, along with the interacting factors, information literacy (referred to erroneously as AI literacy in several places, including the title), and cognitive fatigue. In line with a growing body of previous research, a negative relationship was found between AI use/dependence and critical thinking. However, this negative relationship was 'buffered' (in the author's words) by information literacy, suggesting the latter might be a core skill to focus on in the age of AI as it might inoculate students from the negative effects of AI use on learning.

Tomisu, H., Ueda, J., & Yamanaka, T. (2025). The cognitive mirror: A framework for AI-powered metacognition and self-regulated learning. *Frontiers in Education, 10,* Article 1697554. https://doi.org/10.3389/feduc.2025.1697554

This is a key conceptual paper that proposes a radical design solution to the cognitive offloading problem. It argues for a fundamental shift from the "AI as Oracle" paradigm (which invites passive offloading) to an "AI as Cognitive Mirror". In this model, the AI is engineered as a teachable novice with a pedagogically useful deficit. It feigns confusion, asking clarifying questions that *force* the learner to engage in explanation and reflection (the Protégé Effect). This approach recenters human agency and externalises the learner's thought processes, making it a useful design framework for promoting the effortful work required for durable learning.

Tricot, A., & Sweller, J. (2014). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review, 26*(2), 265-283. https://doi.org/10.1007/s10648-013-9243-1

This theoretical paper synthesises expert-novice research to argue that domain-specific knowledge, not generic skills, drives intellectual expertise. Using Geary's evolutionary framework, the authors distinguish biologically primary skills (evolved, automatically acquired) from secondary knowledge (cultural, requiring instruction). Generic skills are unteachable because they're biologically primary, while domain-specific skills require explicit instruction. The paper explains paradoxical findings across chess, programming, and problem-solving through this lens. This work challenges educational paradigms emphasising "21st century skills" and provides theoretical foundations for content-focused instruction, proving especially relevant as AI handles generic tasks while domain expertise becomes the critical human differentiator.

Tu, Y., Chen, J., & Huang, C. (2025). Empowering personalized learning with generative artificial intelligence: Mechanisms, challenges and pathways. *Frontiers in Digital Education, 2*, 19. https://doi.org/10.1007/s44366-025-0056-9

This technical/theoretical study attempted to plot a realistic course to truly adaptive AI systems for supporting personalised learning. While the authors outline the potential for these systems, their work highlights significant technical and pedagogical barriers to truly personalised learning, even with the power of LLMs added to the mix. This serves as a caution that *current* AI tools are not the sophisticated, adaptive tutors that proponents often describe.

Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education, 21*(1), Article 15. https://doi.org/10.1186/s41239-024-00448-3

This theoretical paper includes arguments for the importance of AI literacy and prompt engineering in education. Time has not been kind to this article. The arguments are outdated, with more recent discussion and research (e.g., Tankelevitch et al., 2024; Walton et al., 2025) deprioritising the kinds of specific technical skills advocated for in this article in favour of more human-focused, metacognitive skills like judgment and verification.

Walton, J., Bearman, M., Crawford, N., Tai, J., & Boud, D. (2025). How university students work on assessment tasks with generative artificial intelligence: Matters of judgement. *Assessment & Evaluation in Higher Education, 50*(5), 705–721. https://doi.org/10.1080/02602938.2025.2570328

A qualitative study looking at how university students use generative AI in assessment tasks. A core theme emerged around the importance of students being able to *judge* their own level of understanding and how this is reflected in the interactions with the technology and, ultimately, in the quality of the work. This study provides the qualitative voice for the metacognitive moderation model, confirming that the learner's own metacognitive monitoring is the critical factor.

Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. *Humanities and Social Sciences Communications, 12*(1), Article 621. https://doi.org/10.1057/s41599-025-04787-y

This meta-analysis of research published between 2022 and 2025 shows that generative AI seems to be beneficial for learning (performance) if used carefully. However, the authors explicitly note that the lack of studies on *higher-order thinking* and learning perceptions (judgements of learning) was a significant gap. The paper seems to have a positive bias because the findings focus on performance, and it implicitly supports the critique that the field is measuring performance, not the deeper learning that other studies find is compromised.

Wang, P., Liu, T., Yang, Y., & Xiang, X. (2025). Optimizing self-regulated learning: A mixed-methods study on GAI's impact on undergraduate task strategies and metacognition. *British Journal of Educational Technology*. Advance online publication. https://doi.org/10.1111/bjet.70018

The mixed-methods study, including an experiment, reported in this paper examined language learning using generative AI in a population of adult learners. The participants benefited from the use of generative AI, including for enhancing their metacognitive processes. However, participants also reported that they were *concerned about becoming over-reliant* on generative AI in their learning. This suggests that the positive effect on metacognition may be moderated by a pre-existing high level of metacognitive awareness.

Wang, R. E., Ribeiro, A. T., Robinson, C. D., Loeb, S., & Demszky, D. (2024). Tutor CoPilot: A human-AI approach for scaling real-time expertise. arXiv preprint arXiv:2410.03017. https://arxiv.org/abs/2410.03017

This preregistered RCT tested a human-AI system providing real-time pedagogical guidance to mathematics tutors. Analysing 4,136 sessions with 782 tutors and 1,787 disadvantaged students, the study found four percentage point improvements in student pass rates (14 points when actively used). Effects were strongest for lower-rated tutors who achieved outcomes comparable to higher-rated controls. The system costs $20 annually versus $3,300+ for traditional professional development, a 165-fold reduction. This first RCT of human-AI collaboration in live education demonstrates viable expertise scaling while maintaining human agency, with particular significance for educational equity.

Wang, T., & Lajoie, S. P. (2023). How does cognitive load interact with self-regulated learning? A dynamic and integrative model. *Educational Psychology Review, 35*(3), Article 69. https://doi.org/10.1007/s10648-023-09794-6

An overview of the relationships between Self-Regulated Learning (SRL) and Cognitive Load Theory (CLT). It makes the point that engaging in SRL *does* create additional cognitive load. However, doing so seems to lead to the development of enhanced learning strategies over time. This paper is foundational, as it explains the *motive* for cognitive offloading: students may rationally choose to bypass the immediate, high-load *cost* of SRL by offloading the task to AI, thereby failing to achieve the long-term *benefit* of schema acquisition.

Weidlich, J., Gašević, D., Drachsler, H., & Kirschner, P. (2025). ChatGPT in education: An effect in search of a cause. *Journal of Computer Assisted Learning, 41*(4), e70105. https://doi.org/10.1111/jcal.70105

This powerful methodological critique argues that the nascent research on generative AI in education (e.g., Deng et al., 2024) is an "effect in search of a cause". The authors argue that claims of effectiveness are largely unfounded, suffering from "fast science" and long-standing conceptual pitfalls. They identify three "non-negotiables" for valid research: (1) a well-defined treatment, (2) a meaningful control group, and (3) valid outcome measures that reflect *durable learning*, not short-term performance. They argue that AI is a tool (a medium), not an instructional method, and that much of the research fails to make this critical distinction, conflating media comparisons with method comparisons.

Willingham, D. T. (2019). *How to teach critical thinking*. NSW Department of Education. https://education.nsw.gov.au/teaching-and-learning/education-for-a-changing-world/thinking-skills/how-to-teach-critical-thinking

This commissioned paper synthesises cognitive science to explain why generic critical thinking instruction fails. Willingham demonstrates that analysis and synthesis mean different things across disciplines, and transfer research shows students fail to recognise when principles apply to new contexts. Knowledge enables critical thinking through problem recognition, working memory chunking, and providing the necessary background for strategy deployment. The four-step framework includes identifying domain-specific skills and content, planning sequential instruction, and revisiting over the years. This work challenges decontextualised skills training, supporting knowledge-rich curricula and proving especially relevant as domain expertise becomes critical when AI handles generic tasks.

Wolf, M. (2016). *Tales of literacy for the 21st century: The literary agenda*. Oxford University Press.

This monograph synthesises neuroscience research to examine how literacy transforms the brain and changes with digital reading. Wolf argues reading requires neural recycling to create circuits enabling deep reading processes (inferential reasoning, critical analysis, empathy development), requiring sustained attention. Digital reading's second revolution produces faster but more superficial processing, reduced attention, and cognitive impatience. Wolf proposes biliteracy, teaching print reading first to establish deep circuits before introducing digital literacy. This work influences literacy instruction debates and provides neuroscientific foundations for concerns about digital media's cognitive impacts, particularly relevant as AI further transforms reading and writing practices.

Xu, X., Qiao, L., Cheng, N., Liu, H., & Zhao, W. (2025). Enhancing self‐regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *British Journal of Educational Technology, 56*(5), 1842–1863. https://doi.org/10.1111/bjet.13599

This paper describes a relatively small quasi-experimental study examining the inclusion of metacognitive prompts to support learning with generative AI. The authors found that the inclusion of metacognitive prompts significantly enhanced learner capability, particularly in terms of self-regulated learning. This study provides a direct empirical solution to the problem identified by Darvishi et al. (2024); it suggests that AI *can* be used to foster SRL, but only if metacognitive supports are *explicitly* designed into the learning environment.

Yang, K., Raković, M., Liang, Z., Yan, L., Zeng, Z., Fan, Y., Gašević, D., & Chen, G. (2025, March 3–7). Modifying AI, enhancing essays: How active engagement with generative AI boosts writing quality. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference (LAK '25)* (pp. 568–578). ACM. https://doi.org/10.1145/3706468.3706544

This Learning Analytics paper provides the *behavioural trace data* to support the goal-driven CLT framework (Kalyuga & Plass, 2025). By analysing a dataset of 1,445 GAI-assisted writing sessions, the authors identified distinct behavioural patterns. Writers who *actively modified* AI-generated text (indicating active engagement and a learn goal) significantly improved essay quality. Conversely, those who *passively accepted* text without changes (indicating a task completion goal) saw a *decrease* in essay quality. This paper provides a crucial, data-driven link between a learner's *actions* (modify vs. accept), their *implied goal* (learn vs. finish), and the *learning outcome*, empirically demonstrating the mechanism of beneficial vs. detrimental offloading.

Yan, L., Greiff, S., Lodge, J. M., & Gašević, D. (2025). Distinguishing performance gains from learning when using generative AI. *Nature Reviews Psychology, 4*(7), 435–436. https://doi.org/10.1038/s44159-025-00467-5

This article directly articulates the central theme of the "performance paradox," arguing that a fundamental distinction must be made between observable performance (which AI boosts) and durable learning (which it may harm). The authors provide a synthesis of the mechanisms, critiquing research that conflates the two (e.g., Deng et al., 2024). They posit that while AI can reduce cognitive load, this "excessively offloading" undermines the active cognitive engagement required for deep encoding and transfer. The paper integrates key findings from the field, linking this paradox to "metacognitive laziness" (Fan et al., 2024) and the paradox of self-efficacy (Zhang & Xu, 2025). It concludes with a methodological call-to-arms, urging researchers to adopt process-oriented assessments like retention and transfer tests to validate learning, a direct critique of the "fast science" (Weidlich et al., 2025) of performance-only studies.

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour, 8*(10), 1839–1850. https://doi.org/10.1038/s41562-024-02004-5

This review outlines the key promises and pitfalls of AI for learning. The promises often centre on the notion of personalised learning and timely support and feedback for students. On the downside, generative AI is notorious for providing incorrect information, and there are numerous moral and ethical issues surrounding these technologies. The authors argue that substantial, rigorous research is needed to understand the impact of these technologies on learning, echoing the critique of Weidlich et al. (2025).

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*(1), Article 28. https://doi.org/10.1186/s40561-024-00316-7

A relatively small systematic review of 14 articles demonstrating that reliance on generative AI can have negative consequences for learners. The key finding is that *efficiency* is one of the core reasons for an over-reliance on AI developing. This review provides the motivational why for detrimental offloading: students are making a metacognitive *choice* to offload (as per Dunn & Risko, 2016) because they are prioritising the goal of task efficiency over the (more effortful) goal of learning.

Zhang, L., & Xu, J. (2025). The paradox of self-efficacy and technological dependence: Unraveling generative AI's impact on university students' task completion. *The Internet and Higher Education, 65*, 100978. https://doi.org/10.1016/j.iheduc.2024.100978

This study introduces a crucial psychological nuance to the cognitive offloading debate: the "paradox of self-efficacy." The authors found that while AI use *increased* students' task-related self-efficacy, it also *intensified* their technological dependence. This provides a powerful explanation for *why* students choose to offload, even to their own detriment. The AI-induced feeling of competence ("fluency") acts as a misleading metacognitive cue, signalling that deep engagement is unnecessary. This paper provides the why for the metacognitive laziness described by Fan et al. (2024) and directly complements the findings of Lee et al. (2025).

For more information about this publication, please email edtechnetwork@uts.edu.au.

Centre for Social Justice and Inclusion

Network for Quality Digital Education