Review Article

UDC: 159.9:004.8

Received: March 14, 2026.
Revised: May 04, 2026.
Accepted: May 06, 2026.

10.23947/2334-8496-2026-14-1-115-124

Check for updates

Artificial Intelligence in Applied Cognitive Psychology: A Commentary

Benjamin T. Sharpe[1,2], *Mathews Rod*[3], *George Horne*[1*]

[1]Faculty of Social Science and Humanities, University of Chichester, Chichester, UK,
e-mail: b.sharpe@chi.ac.uk, g.horne@chi.ac.uk
[2]Faculty of Health, Southern Cross University, Lismore, Australia
[3]Department of Information and Learning Technologies, University of Chichester, Chichester, UK,
e-mail: R.matthews@chi.ac.uk

Abstract: AI integration in applied cognitive psychology demands critical evaluation beyond efficiency metrics. Despite widespread institutional adoption, emerging research reveals concerning patterns including high hallucination rates, deteriorating retention with prolonged exposure, and a consistent tendency to support surface-level task completion at the expense of deeper cognitive processing. These findings align with established principles regarding desirable difficulties, metacognitive monitoring, skill acquisition, and vigilance, suggesting that applications prioritising task completion over cognitive development risk undermining the adaptive expertise essential for complex professional contexts. Methodological weaknesses in existing research, including brief interventions, inadequate control comparisons, and reliance on satisfaction measures, further constrain confident conclusions. Nonetheless, several domains including cognitive accessibility, rehabilitation, vigilance, and adaptive tutoring represent areas of genuine promise where AI's architecture may complement rather than conflict with established cognitive science. This commentary synthesises emerging evidence, examines methodological limitations, proposes research priorities for responsible integration, and reflects on where cautious optimism is warranted.

Keywords: *Artificial Intelligence, Applied Cognitive Psychology, Expertise Development.*

Introduction

The rapid proliferation of generative artificial intelligence systems (Kasneci et al., 2023) presents applied cognitive psychology with methodological opportunities and conceptual challenges across educational, occupational, clinical, and research domains. This commentary examines AI integration through applied cognitive psychology's evidence-based frameworks, focusing specifically on generative AI systems (large language models such as ChatGPT) and AI-enhanced educational tools currently being adopted across institutions. Amidst technological enthusiasm and institutional pressure for AI adoption, applied cognitive psychology can provide sceptical, evidence-based analysis.

Institutional pressures for AI adoption manifest through administrative mandates for technology integration, competitive positioning amongst peer institutions, and commercial partnerships with technology providers, often proceeding without rigorous evaluation of cognitive outcomes (Williamson, 2021). These commercial relationships create conflicts between vendors' financial interests in widespread adoption and cognitive psychology's scientific obligation to assess genuine educational and professional benefits. Technology companies invest substantial resources in marketing AI tools to educational institutions and professional organisations, often emphasising efficiency gains whilst downplaying or ignoring cognitive development outcomes (Jobin et al., 2019). When universities and organisations adopt AI systems through commercial contracts before rigorous empirical evaluation, they prioritise market positioning over evidence-based decision-making, whilst also placing at risk the statutory privacy rights of students, staff, and service users whose data enters commercial systems without adequate scrutiny.

*Corresponding author: g.horne@chi.ac.uk

Applied cognitive psychology must resist becoming complicit in these premature adoptions driven by commercial interests rather than demonstrated cognitive benefits. The field's credibility depends on maintaining independence from technology industry narratives and insisting on rigorous empirical evaluation before endorsing widespread AI integration. As a discipline concerned with understanding and enhancing human cognitive functioning across domains including memory, attention, learning, decision-making, vigilance, and metacognition, we must evaluate AI through established cognitive science frameworks whilst reconsidering foundational assumptions about human-AI interaction.

Current practice reveals troubling disconnection between evidence and implementation: institutions adopt AI based on efficiency gains whilst ignoring cognitive development outcomes; organisations implement decision-support without assessing impacts on professional judgement or situational awareness; technology companies promote tools without addressing skill atrophy or dependency. This commentary synthesises emerging empirical evidence, examining four questions: (1) What do systematic investigations reveal about AI's cognitive impacts? (2) How do findings relate to established principles? (3) What methodological limitations constrain conclusions? (4) What research priorities should guide responsible integration?

Empirical Evidence on AI in Cognitive Psychology Contexts

Recent meta-analytic evidence provides initial empirical grounding for understanding AI's impact on learning outcomes. Synthesis of 51 studies revealed that AI systems demonstrate large positive effects on learning performance ($g = 0.867$) but substantially smaller effects on 'deeper cognitive processing', more specifically, a moderately positive effect on learning perception ($g = 0.456$) and on higher-order thinking ($g = 0.457$; Wang and Fan, 2025). This differential impact pattern aligns with cognitive psychology's longstanding distinction between surface and deep learning (Marton and Säljö, 1976; Yang et al., 2024), suggesting that AI may facilitate utilisation of factual knowledge whilst neglecting deeper understanding, learning, and critical reflection which would encourage long-term professional development (Li et al., 2024; Marshall et al., 2022). AI usage could be a cognitive crutch (Barcaui, 2025), rather than scaffolding deeper learning. Of particular concern is the risk that users operating at a surface level of engagement may uncritically accept AI-generated content, including hallucinated claims, as factual, with this risk amplifying in proportion to the consequentiality of the decision or claim involved. It is worth noting that the meta-analysis itself has attracted methodological scrutiny, with critics noting that the included studies were not assessed for peer-review status, randomisation, or statistical power, a limitation that warrants cautious interpretation of the effect size estimates.

This pattern holds concerning implications beyond educational contexts: if AI similarly facilitates surface-level task completion in professional settings whilst undermining development of deep domain expertise, organisations may cultivate workforces capable of executing routine procedures using AI assistance yet lacking the adaptive expertise required when AI systems fail, prove inapplicable, or when novel situations demand flexible problem-solving. This concern maps onto Argyris and Schön's (1978) distinction between single-loop and double-loop learning: AI may reinforce single-loop responses, in which errors are corrected within existing assumptions, whilst impeding the deeper, paradigm-questioning reflection characteristic of double-loop learning that is essential for adaptive professional expertise.

This pattern gains theoretical significance when considered alongside cognitive load theory (Sweller, 1988; Sweller et al., 1998) and research on desirable difficulties in learning (Bjork, 1994; Bjork and Bjork, 2011, 2020). If AI systems complete cognitive operations that learners or professionals would otherwise perform, generating explanations, constructing connections, monitoring comprehension, then reduced cognitive effort may paradoxically undermine the very processing that supports durable learning and transfer. Yang et al. (2024) distinguish surface from deep learning approaches with generative AI, suggesting that effectiveness depends critically on pedagogical or operational integration rather than mere AI tool availability.

Longitudinal evidence further complicates simple assessments of AI's value for cognitive performance and development. Akgun and Toker (2024) documented that whilst pretesting before AI use enhanced retention, consistent with testing effect research (Roediger and Karpicke, 2006), prolonged AI exposure led to memory decline in their sample of 73 undergraduates. This finding suggests potential time-dependent effects wherein initial benefits may give way to dependency or reduced cognitive engagement, paralleling concerns about calculators and GPS navigation systems affecting mathematical and spatial cognition respectively. Recent research on cognitive offloading to external tools demonstrates that whilst offloading can improve immediate task performance, it frequently results in reduced memory for offloaded information (Grinschgl et al., 2021). Experimental evidence shows that increasing cognitive offloading improves performance but diminishes subsequent memory, with participants showing lower recall for offloaded material compared to internally processed information. Research on cognitive offloading reveals that whilst offloading can temporarily free cognitive resources, it frequently results in skill atrophy when individuals become dependent on tools for operations they previously performed internally (Risko and Gilbert, 2016; Gilbert et al., 2023). Grinschgl et al. (2021) demonstrated that reducing offloading costs led to increased offloading behaviour alongside improved immediate performance but significantly diminished memory in subsequent tests. Ward et al. (2017) found that mere smartphone presence reduces available cognitive capacity even when devices remain unused. These findings suggest AI availability may create similar cognitive costs through reduced encoding effort and increased dependency. In occupational contexts, such patterns are particularly concerning emergency responders, medical professionals, or system operators who rely on AI decision-support may experience degraded performance when AI systems malfunction or when situations demand rapid independent judgement without technological assistance.

Research on AI's impact on critical thinking reveals nuanced patterns requiring careful interpretation. Essel et al.'s (2024) investigation with undergraduates identified improvements in question formulation and information analysis alongside concerning patterns of passive acceptance of AI-generated responses. Similarly, Ododo et al. (2024) documented threats to critical thinking when students uncritically accepted AI information without analytical engagement or scrutiny. These findings resonate with decades of cognitive psychology research emphasising that critical thinking develops through active engagement with challenging material rather than passive consumption of information (Chi, 2009). In professional contexts requiring complex judgement, management decision-making, clinical assessment, safety-critical evaluation, uncritical acceptance of AI recommendations without independent verification represents significant risk.

From a metacognitive perspective, AI systems' inability to model metacognitive monitoring represents a fundamental limitation (Exintaris et al., 2023). Metacognition, encompassing monitoring of one's understanding, recognition of knowledge boundaries, and strategic regulation of cognitive processes, has been established as central to effective learning and expert performance (Flavell, 1979; Schraw and Dennison, 1994). Research on metacognitive calibration demonstrates that professionals who accurately assess their understanding show superior outcomes compared to those with poor confidence-accuracy correspondence (Baars et al., 2014). The illusion of knowing, wherein individuals feel confident despite incomplete understanding, becomes particularly problematic when AI systems provide immediate answers without requiring users to evaluate their own knowledge state first. When AI systems confidently present inaccurate information without appropriate uncertainty calibration, they model precisely the metacognitive dysfunction that cognitive psychology seeks to prevent.

Systematic investigation of AI accuracy reveals substantial concerns for cognitive psychology applications. A recent systematic review documented hallucination rates ranging from 28-91% across AI models, with considerable variance depending on task complexity and domain (Chelli et al., 2024). Specifically, these rates were 39.6% for GPT-3.5, 28.6% for GPT-4, and 91.4% for Bard, indicating substantial variation in reliability across models and making the upper bound largely attributable to an earlier-generation system. Examination of ChatGPT citations found 32.3% were fabricated (MacDonald, 2023), whilst broader systematic reviews indicate that whilst AI can reduce literature review workload by 60-65%, precision varies dramatically from 4.6% to 88% depending on task characteristics (Chelli et al., 2024). These accuracy limitations present challenges for applied cognitive psychology, where interventions, organisational recommendations, and clinical applications must be grounded in accurate representation of research evidence. When AI systems conflate distinct theoretical constructs, misattribute empirical findings, or fabricate research that appears plausible to non-experts, they undermine the evidence-based foundation essential to responsible professional practice, and raise fundamental questions about the conditions under which AI-derived output can be treated as information rather than raw data. Without robust supervisory frameworks and reflexive evaluation practices, practitioners risk ascribing evidential weight to content that is, at best, unverified and, at worst, systematically misleading. In safety-critical domains, incorrect AI recommendations regarding vigilance maintenance, workload management, or decision-support system design could compromise operational safety, particularly in settings such as lifeguarding where failure to detect a hazard can lead to permanent injury or death (Sharpe et al., 2024; Vansteenkiste et al., 2025).

Beyond individual cognitive impacts, AI integration raises substantial equity concerns that applied cognitive psychology must address. Works on digital divides demonstrates that technology access follows existing socioeconomic stratification patterns (Van Dijk, 2020). Critical discussions of AI in education consistently identify digital equity as a major challenge, with AI tools requiring subscription costs, high-bandwidth internet, or advanced digital literacy potentially exacerbating rather than reducing educational and professional inequalities (Rottner et al., 2025). Individuals from under-resourced communities may lack access to AI tools their more privileged peers utilise, potentially widening achievement gaps and limiting social mobility (Mostafa, 2025). Furthermore, AI systems trained predominantly on data from Western, educated, industrialised, rich, and democratic populations may perform poorly for culturally and linguistically diverse users, embedding bias into ostensibly objective technological systems (Rottner et al., 2025). The historical record on which LLMs are predominantly trained reflects not only geographic and economic stratification but also a substantial gender skew, given that recorded history, published scholarship, and digitised public discourse have been disproportionately produced by and about men, creating conditions in which gender bias may be structurally encoded rather than incidental. Applied cognitive psychology's commitment to understanding human cognition across diverse populations demands critical examination of who benefits from AI integration and who faces exclusion or disadvantage. Premature adoption without addressing equity concerns risks institutionalising technological advantages for already-privileged groups whilst disadvantaging those most needing educational and professional support.

Methodological Limitations of Current AI Research

Critical examination of existing AI research reveals substantial methodological weaknesses that constrain confident conclusions about AI's cognitive impacts. Many studies employ brief interventions, typically single sessions or fewer than four weeks, insufficient for assessing impacts on skill development or knowledge consolidation. These brief timeframes may capture initial novelty effects or short-term performance changes whilst missing longer-term patterns including skill atrophy, dependency development, or metacognitive deterioration. In occupational contexts where expertise develops over months or years through extensive deliberate practice, such abbreviated studies cannot adequately assess AI's impact on professional competence.

Control group selection in AI research frequently proves inadequate for establishing AI's unique contribution. Many studies compare AI use against no intervention rather than against evidence-based alternatives, creating situations where any active intervention would likely demonstrate superiority over passive control conditions. Rigorous evaluation requires comparing AI against validated approaches: worked examples, practice testing, or spaced repetition in educational contexts; established decision-support systems or expert consultation in occupational settings. Without such comparisons, apparent AI benefits may simply reflect general support rather than AI-specific advantages.

Outcome assessment represents another critical limitation. Many AI studies rely on self-reported satisfaction, perceived usefulness, or intention to continue use rather than objective cognitive performance measures. Whilst subjective experience matters, it proves insufficient for evaluating cognitive impact. Research consistently demonstrates that learners poorly judge which study strategies produce durable learning (Koriat and Bjork, 2005; Soderstrom and Bjork, 2015), often preferring approaches that create illusions of competence over those supporting genuine mastery. Similarly, AI users may report high satisfaction whilst experiencing metacognitive impairment, dependency development, or skill atrophy undetected by satisfaction measures. Rigorous evaluation demands validated cognitive assessments including retention tests, transfer tasks, and metacognitive calibration measures administered at delayed timepoints.

Sample characteristics further limit generalisability. Most AI research employs undergraduate convenience samples, creating uncertainty about findings' applicability to professional contexts requiring extensive domain expertise. University students and domain experts differ substantially in prior knowledge, metacognitive sophistication, and task approach strategies, factors that may moderate AI's cognitive impact. A medical consultant's interaction with AI decision-support likely differs fundamentally from an undergraduate's use of ChatGPT for essay writing. Extrapolating from student samples to professional contexts risks overlooking critical expertise-related moderators.

Publication bias represents a pervasive concern across AI research. The methodological enthusiasm surrounding AI, combined with commercial interests and institutional pressures, creates conditions favouring positive findings. Studies documenting AI limitations, null effects, or negative outcomes may face publication barriers, potentially biasing the evidence base towards optimistic conclusions. This systematic bias undermines meta-analytic synthesis and evidence-based decision-making.

Theoretical Foundations and Cognitive Science Principles

Established cognitive science principles provide frameworks for understanding AI's impacts and limitations. The concept of desirable difficulties (Bjork, 1994; Bjork and Bjork, 2011, 2020) illustrates that conditions impeding immediate performance often enhance long-term retention and transfer. Strategies including generation, spacing, interleaving, and variation create processing demands that, whilst temporarily reducing performance, strengthen memory and promote flexible knowledge application. If AI systems eliminate these beneficial difficulties, providing immediate answers, removing retrieval demands, minimising error correction, they may inadvertently undermine the cognitive processing essential for durable learning.

The testing effect (Roediger and Karpicke, 2006; Rowland, 2014) demonstrates that retrieval practice produces superior retention compared to repeated study, even when retrieval proves difficult and error-prone. AI systems that provide immediate answers without requiring retrieval bypass this mechanism, potentially reducing long-term retention. Similarly, research on worked examples and problem-solving (Sweller and Cooper, 1985; Kalyuga et al., 2003) reveals complex interactions between learner expertise and instructional support. Whilst worked examples benefit novices, they impede expert learning, a pattern termed the expertise reversal effect. AI assistance may similarly help novices whilst limiting expert development if it eliminates the problem-solving practice essential for building adaptive expertise.

Metacognitive monitoring (Flavell, 1979; Koriat, 2007) enables effective learning through accurate assessment of understanding and strategic regulation of study behaviour. When AI systems provide answers without prompting metacognitive evaluation, users may develop inflated confidence despite incomplete understanding. This metacognitive impairment extends beyond education: professionals using AI decision-support may experience reduced calibration between confidence and accuracy, potentially leading to overreliance on AI recommendations and diminished independent judgement.

Research on skill acquisition (Ericsson et al., 1993) emphasises that expertise requires extensive deliberate practice with immediate feedback on challenging tasks within one's domain. If AI completes tasks that would otherwise constitute deliberate practice, users may never develop the skills essential for independent performance. This concern proves particularly salient in professional contexts where AI may facilitate task completion whilst preventing the practice essential for building adaptive expertise.

Vigilance research (Warm et al., 2008; Hancock, 2013) reveals that sustained attention degrades over time, particularly for rare events in monotonous contexts. AI systems designed to support vigilance may inadvertently reduce operators' attentional engagement, leading to greater performance decrements when AI fails or proves unavailable. In safety-critical domains including lifeguarding, aviation, and medical monitoring, such vigilance decrements could compromise safety (Sharpe et al., 2026).

AI integration raises fundamental epistemological concerns for cognitive psychology's scientific foundations. Traditional scientific understanding emphasises transparent, reproducible reasoning processes that can be evaluated, critiqued, and refined. AI systems, particularly large language models, function as "black boxes" whose decision-making processes remain opaque even to their developers (Rudin, 2019; Ahmed et al., 2022). This opacity fundamentally conflicts with scientific principles of explanatory understanding. When AI provides correct answers through inscrutable processes, it shifts emphasis from explanatory knowledge to predictive performance, a transition that may undermine scientific literacy and critical evaluation skills central to cognitive psychology training. Recent analyses of explainable AI methods reveal that whilst techniques exist to increase transparency, many create false impressions of understanding through inconsistent or misleading explanations (Durán and Jongsma, 2021). Professionals who rely on AI-generated insights without understanding underlying mechanisms may struggle to 1.) recognise when AI fails, 2.) generalise knowledge to novel contexts, or 3.) explain their reasoning to colleagues and clients. It is also worth noting that failure in AI systems is not binary: an unsubstantiated, ethically compromised, or subtly misleading output represents a form of failure even when superficially plausible, and this risk is compounded by the feedback dynamics of large language model training, wherein widely shared or repeatedly accessed hallucinations may be inadvertently reinforced as training data for subsequent model iterations, progressively entrenching inaccurate information within the knowledge base on which the field increasingly relies. This epistemological shift from transparent understanding to opaque prediction represents a fundamental challenge to applied cognitive psychology's commitment to evidence-based practice grounded in comprehensible theoretical frameworks. If practitioners cannot explain how conclusions were reached or assess the validity of AI-generated recommendations, the field risks losing the explanatory coherence that distinguishes professional expertise from algorithmic output.

Research Priorities for Responsible AI Integration

Applied cognitive psychology must establish rigorous research priorities to guide responsible AI integration. Longitudinal studies examining AI's long-term cognitive impacts across diverse populations and contexts are essential. Such research should employ validated cognitive assessments administered at delayed timepoints to capture retention, transfer, and metacognitive effects invisible in immediate performance measures. Studies should examine expertise development trajectories, assessing whether AI use facilitates or impedes progression from novice to expert performance. Comparative studies must evaluate AI against evidence-based alternatives rather than passive control conditions. In educational contexts, comparisons should include worked examples, practice testing, spaced repetition, and other validated learning strategies. In occupational settings, evaluations should compare AI against established decision-support systems, expert consultation, and traditional training approaches. Such comparisons will clarify whether AI provides unique benefits or merely replicates existing effective practices.

Research must examine individual differences moderating AI's cognitive impact. Expertise, metacognitive sophistication, learning strategies, and domain knowledge likely influence how people interact with AI and whether interactions prove beneficial or detrimental. Identifying these moderators will enable targeted recommendations about when, for whom, and under what conditions AI proves appropriate. Ecological validity demands investigation of AI in authentic professional contexts with domain experts performing realistic tasks. University student performance on artificial laboratory tasks may poorly predict professional use of AI for complex judgement, diagnosis, or safety-critical decision-making. Field studies in medical, educational, legal, and engineering contexts will provide essential evidence about real-world impacts.

Research should examine AI's effects on collaborative work and team performance. Most current research focuses on individual users, yet many professional contexts involve teamwork. A conceptually important distinction concerns whether AI functions as an independent contributor to group cognition, a kind of synthetic team member whose outputs influence shared mental models, or as a parallel tool available individually to each member, with substantially different implications for group dynamics, accountability, and collective bias depending on the answer, particularly where a single proprietary system shapes the informational environment of all members simultaneously. AI's impact on communication, coordination, shared mental models, and collective performance remains largely unexplored. Mechanism-focused research should identify the cognitive processes through which AI affects learning, expertise development, and professional performance. Understanding whether AI impacts primarily encoding, retrieval, metacognition, attention, or problem-solving strategies will inform more effective AI design and implementation. Intervention research should examine pedagogical and operational strategies for maximising AI benefits whilst minimising cognitive costs. Such research might explore optimal timing of AI introduction, appropriate scaffolding strategies, methods for maintaining metacognitive engagement, and approaches for preventing unhealthy dependency or skill atrophy.

Beyond empirical research, applied cognitive psychology must engage with governance and regulatory frameworks for AI in professional practice. Currently, minimal professional guidelines exist regarding appropriate AI use in psychological assessment, intervention design, clinical decision-making, or research synthesis (Meskó and Topol, 2023). Analysis of 793 state-level AI bills in the United States identified only 28 explicitly referencing mental health applications, with substantial gaps in professional oversight, harm prevention protocols, and data governance (Shumate et al., 2025). Meta-analysis of 200 global AI governance documents reveals widespread calls for accountability, transparency, and fairness, yet most guidelines remain non-binding recommendations rather than enforceable standards (Jobin et al., 2019). This regulatory vacuum creates risks including inappropriate delegation of professional judgement to AI systems, inadequate informed consent when AI contributes to clinical decisions, and unclear accountability when AI-assisted interventions produce adverse outcomes. Professional psychology organisations must develop evidence-based guidelines specifying when AI proves appropriate, what safeguards protect clients and research participants, and how practitioners maintain competence and professional responsibility, legal obligations under applicable data protection and privacy law, and adherence to the ethical principles enshrined in professional codes of practice when utilising AI tools (Martinez-Martin et al., 2020). Such guidelines should address mandatory training requirements for AI use in professional contexts, standards for evaluating AI system validity and reliability, and protocols for monitoring AI impacts on client outcomes. Applied cognitive psychology must also contribute to broader regulatory discussions, ensuring that AI deployment in educational, clinical, and occupational settings reflects genuine understanding of cognitive impacts rather than uncritical technological enthusiasm. Without proactive engagement in governance development, the field risks having AI regulation imposed by policymakers lacking psychological expertise, potentially mandating practices inconsistent with cognitive science evidence.

The Authors Optimism

It is worth acknowledging upfront that the applications highlighted in this subsection are inevitably shaped by the authors' own research interests and should be read accordingly. With that caveat noted, genuine promise does exist alongside the documented limitations and identifying where AI's cognitive footprint may differ substantially from the educational and professional contexts critiqued above is not a rhetorical gesture towards balance but a precondition for directing the research agenda towards the most consequential questions. One underappreciated area of genuine promise concerns AI's potential to support its own quality assurance. There is a credible case that AI systems, embedded within appropriately designed supervisory frameworks, could be leveraged to detect the probabilistic signatures of hallucinated content, flag claims that fall outside the evidential range of reliable training data, and generate calibrated uncertainty statements that prompt rather than suppress critical evaluation. Such an application would not resolve the foundational epistemological concerns raised elsewhere in this commentary, but it would represent a meaningful step towards reflexive AI deployment, one in which the technology participates in the governance of its own outputs rather than leaving that burden entirely to users who may lack the domain knowledge to identify what they are being given.

Cognitive accessibility and neurodivergent populations represent one of the most compelling and underexamined areas of opportunity. For individuals with dyslexia, ADHD, autism spectrum conditions, or acquired cognitive impairments, AI-driven tools including adaptive text reformatting, real-time summarisation, and executive-function scaffolding offer the potential to reduce extrinsic cognitive load in ways that may genuinely support rather than bypass cognitive engagement. The desirable difficulties framework may apply quite differently when the difficulty being reduced is a structural barrier rather than a productive cognitive challenge, and applied cognitive psychology has the theoretical vocabulary to articulate this distinction precisely, yet this literature has developed largely in isolation from mainstream AI cognition research. Further, AI-personalised cognitive training and rehabilitation programmes offer a second area of considerable promise. Unlike general-purpose LLMs, systems designed for dementia prevention, acquired brain injury rehabilitation, or age-related cognitive decline are structured explicitly to exercise cognitive mechanisms rather than offload them, making the dependency and skill atrophy concerns raised throughout this commentary considerably less salient. The capacity of AI to personalise difficulty trajectories, monitor engagement, and deliver intervention at scale addresses a genuine limitation of traditional programme delivery.

A third underexplored application lies in vigilance augmentation within safety-critical domains. The distinction between AI as attentional substitute and AI as perceptual amplifier maps directly onto existing cognitive theory around levels of automation and situation awareness, representing a domain where vigilance research could yield practically significant design principles across lifeguarding, air traffic control, and intensive care monitoring. Finally, adaptive intelligent tutoring systems that embed retrieval practice, spacing, and metacognitive prompting represent an application where AI architecture can be aligned with established learning science rather than placed in tension with it. These systems predate the current wave of generative AI and carry a more robust empirical foundation. The prospect of realising such applications responsibly is itself a compelling reason for the field to invest in exactly the methodological standards and governance frameworks this commentary advocates.

Conclusion

AI integration in applied cognitive psychology requires evidence-based evaluation of cognitive impacts rather than uncritical adoption driven by technological enthusiasm or institutional pressures. Current evidence reveals differential effects, with AI supporting surface learning more effectively than deeper cognitive processing. Substantial methodological limitations constrain confident conclusions about AI's long-term impacts on expertise development, metacognition, and professional performance. Established cognitive science principles regarding desirable difficulties, testing effects, metacognitive monitoring, skill acquisition, and vigilance provide frameworks for understanding both AI's potential benefits and risks. Critical analysis reveals concerning equity implications, epistemological challenges, and governance gaps that demand urgent attention.

Responsible integration demands systematic investigation addressing current methodological weaknesses through longitudinal designs, appropriate control comparisons, objective cognitive assessments, diverse samples, and authentic professional contexts. Applied cognitive psychology must resist institutional pressures for premature adoption, instead providing sceptical, evidence-based analysis of when, for whom, under what conditions, and through what mechanisms AI enhances or undermines cognitive functioning. Only through rigorous empirical investigation and proactive governance engagement can the field fulfil its responsibility to promote genuine cognitive development rather than mere task completion in an increasingly AI-integrated world.

References

Ahmed, I., Jeon, G., & Piccialli, F. (2022). From artificial intelligence to explainable artificial intelligence in industry 4.0: A survey on what, how, and where. *IEEE Transactions on Industrial Informatics, 18*(8), 5031–5042. https://doi.org/10.1109/TII.2022.3146552

Argyris, C., & Schön, D. A. (1978). *Organizational learning: A theory of action perspective.* Addison-Wesley.

Akgun, S., & Toker, S. (2024). Evaluating the Effect of Pretesting with Conversational AI on Retention of Needed Information. *ArXiv.* https://doi.org/10.48550/arXiv.2412.13487

Baars, M., Vink, S., van Gog, T., de Bruin, A., & Paas, F. (2014). Effects of training self-assessment and using assessment standards on retrospective and prospective monitoring of problem solving. *Learning and Instruction, 33,* 92–107. https://doi.org/10.1016/j.learninstruc.2014.04.004

Barcaui, A. (2025). ChatGPT as a cognitive crutch: Evidence from a randomized controlled trial on knowledge retention. *Social Sciences & Humanities Open*, *12*, 102287. https://doi.org/10.1016/j.ssaho.2025.102287

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), Psychology and the real world: Essays illustrating fundamental contributions to society, 56–64. Worth Publishers.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), Metacognition: Knowing about knowing, 185–205. MIT Press.

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition, 9*(4), 475–479. https://doi.org/10.1016/j.jarmac.2020.09.003

Chelli, M., Descamps, J., Lavoué, V., Trojani, C., Azar, M., Deckert, M., Raynier, J. L., Clowez, G., Boileau, P., & Ruetsch-Chelli, C. (2024). Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: Comparative analysis. *Journal of Medical Internet Research, 26*, e53164. https://doi.org/10.2196/53164

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

Durán, J. M., & Jongsma, K. R. (2021). Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. *Journal of Medical Ethics, 47*(5), 329–335. https://doi.org/10.1136/medethics-2020-106820

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363–406. https://doi.org/10.1037/0033-295X.100.3.363

Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence, 6*, 100198. https://doi.org/10.1016/j.caeai.2023.100198

Exintaris, B., Karunaratne, N., & Yuriev, E. (2023). Metacognition and critical thinking: Using ChatGPT-generated responses as prompts for critique in a problem-solving workshop (SMARTCHEMPer). *Journal of Chemical Education, 100*(8), 2972–2980. https://doi.org/10.1021/acs.jchemed.3c00481

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Shumate, J. N., Rozenblit, E., Flathers, M., Larrauri, C. A., Hau, C., Xia, W., ... & Torous, J. (2025). Governing AI in mental health: 50-state legislative review. *JMIR Mental Health*, *12*, e80739.

Gilbert, S. J., Boldt, A., Sachdeva, C., Scarampi, C., & Tsai, P.-C. (2023). Outsourcing memory to external tools: A review of 'intention offloading'. *Psychonomic Bulletin & Review, 30*(1), 60-76. https://doi.org/10.3758/s13423-022-02139-4

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

Hancock, P. A. (2013). In search of vigilance: the problem of iatrogenically created psychological phenomena. *American Psychologist, 68*(2), 97-109. https://doi.org/10.1037/a0030214

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23–31. https://doi.org/10.4018/978-1-60566-048-6.ch003

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 187-194. https://doi.org/10.1037/0278-7393.31.2.187

Koriat, A., & Helstrup, T. (2007). Metacognitive aspects of memory. In *Everyday memory* (pp. 251-274). Psychology Press.

Li, T., Ji, Y., & Zhan, Z. (2024). Expert or machine? Comparing the effect of pairing student teacher with in-service teacher and ChatGPT on their critical thinking, learning performance, and cognitive load in an integrated-STEM course. *Asia Pacific Journal of Education*, *44*(1), 45-60. https://doi.org/10.1080/02188791.2024.2305163

MacDonald, R. (2023). Dude, where's my citations? ChatGPT's hallucination of citations. Mind Pad: The Canadian Psychological Association's Magazine for Psychology Students and Instructors, Winter 2023. https://cpa.ca/docs/File/Students/MindPad/MindPad_Winter2023.pdf

Marshall, T., Keville, S., Cain, A., & Adler, J. R. (2022). Facilitating reflection: a review and synthesis of the factors enabling effective facilitation of reflective practice. *Reflective Practice*, *23*(4), 483-496. https://doi.org/10.1080/14623943.2022.2064444

Martinez-Martin, N., Luo, Z., Kaushal, A., Adeli, E., Haque, A., Kelly, S. S., ... Char, D. S. (2020). Ethical issues in using ambient intelligence in health-care settings. *The Lancet Digital Health, 2*(2), e115–e123. https://doi.org/10.1016/S2589-7500(20)30275-2

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I, Outcome and process. *British Journal of Educational Psychology, 46*(1), 4–11. https://doi.org/10.1111/j.2044-8279.1976.tb02980.x

Meskó, B., & Topol, E. J. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ digital medicine, 6*(1), 120. https://doi.org/10.1038/s41746-023-00873-0

Mostafa, M. A. M. (2025). Bridging the digital divide: AI, VR, and AR for equitable K-12 education. *SSRN Electronic Journal*. https://dx.doi.org/10.2139/ssrn.5124551

Ododo, E. P., Iniobong, U. B., Udoessien, A. I., Ukpe, I. U., & James, O. D. (2024). Artificial intelligence in the classroom: Perceived challenges to vocational education student retention and critical thinking in tertiary institutions. *American Journal of Interdisciplinary Innovative Research*, *6*(9), 30-39. https://doi.org/10.37547/tajiir/Volume06Issue09-05

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

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rottner, R., Porter, L., Bock, J., Jannone, J., Senerchia, R. W., Ward, J., & Whittinghill, J. (2025). AI and the Digital Divide. In Corbeil, J. R., & Corbeil, M. E. (Eds). *Teaching and Learning in the Age of Generative AI* (pp. 309-331). Routledge.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological bulletin, 140*(6), 1432-1463. https://doi.org/10.1037/a0037559

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence, 1*(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*(4), 460–475. https://doi.org/10.1006/ceps.1994.1033

Sharpe, B. T., Cotten, E. L., & Sivyer, L. R. (2026). The dangers of occupational vigilance: a scoping review. *Safety and Reli-*

ability, 1–23. https://doi.org/10.1080/09617353.2025.2607194

Sharpe, B. T., Smith, M. S., Williams, S. C., Hampshire, A., Balaet, M., Trender, W., ... Smith, J. (2024). Cognition and lifeguard detection performance. *Applied Cognitive Psychology, 38*(1), e4139. https://doi.org/10.1002/acp.4139

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

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59–89. https://doi.org/10.1207/s1532690xci0201_3

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, *10*(3), 251-296. https://doi.org/10.1023/A:1022193728205

Van Dijk, J. A. G. M. (2020). *The digital divide*. Polity Press.

Vansteenkiste, P., Bourgois, J. G., & Lenoir, M. (2025). Baywatch in the laboratory, Differences in visual surveillance between lifeguards and non-lifeguards. *Applied Cognitive Psychology, 39*(5), e70110. https://doi.org/10.1002/acp.70110

Wang, X., & Fan, Y. (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*, 621. https://doi.org/10.1057/s41599-025-04787-y

Ward, A. F., Duke, K., Gneezy, A., & Bos, M. W. (2017). Brain drain: The mere presence of one's own smartphone reduces available cognitive capacity. *Journal of the Association for Consumer Research, 2(*2), 140–154. https://doi.org/10.1086/691462

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. Human Factors, 50(3), 433–441. https://doi.org/10.1518/001872008X312152

Williamson, B. (2021). Meta-edtech: Digital education governance, data science and algorithmic power. *Learning, Media and Technology, 46*(1), 1–2. https://doi.org/10.1080/17439884.2021.1876089

Yang, Y., Luo, J., Yang, M., & Chen, J. (2024). From surface to deep learning approaches with generative AI in higher education: An analytical framework of student agency. *Studies in Higher Education, 49*(5), 817–830. https://doi.org/10.1080/03075079.2024.2327003