Scientific Research Publishing

Open Journal of Social Sciences, 2026, 14(3), 251-267
https://www.scirp.org/journal/jss
ISSN Online: 2327-5960
ISSN Print: 2327-5952

Beyond Efficiency: A Qualitative Exploration of Human Agency, Epistemic Vigilance, and Cognitive Boundaries in Human-AI Interaction

Mohammed Tayyab Khan[1], Jonathan Ee[2], Shilpi Tripathi[3]

[1]School of Social Sciences, London Metropolitan University, London, UK
[2]School of Psychology, London Metropolitan University, London, UK
[3]Independent Researcher, Singapore
Email: tripathi888@gmail.com

How to cite this paper: Khan, M. T., Ee, J., & Tripathi, S. (2026). Beyond Efficiency: A Qualitative Exploration of Human Agency, Epistemic Vigilance, and Cognitive Boundaries in Human-AI Interaction. *Open Journal of Social Sciences, 14,* 251-267.
https://doi.org/10.4236/jss.2026.143015

Received: February 4, 2026
Accepted: March 14, 2026
Published: March 17, 2026

Abstract

As Artificial Intelligence (AI) becomes more common in professional and psychological contexts, concerns about human autonomy and epistemic vigilance have grown. Questions around the changing boundaries of cognitive responsibility are intensifying. This qualitative study explores how professionals in psychology, leadership, and technology negotiate agency, cognitive effort, and scrutiny when using AI. Through reflexive thematic analysis of twelve semi-structured interviews, two superordinate themes emerged: 1) Human Agency and Epistemic Vigilance, and 2) Cognitive Boundaries and the Recalibration of Mental Effort. Participants described AI as a powerful accelerator of cognitive efficiency, but noted its potential to disrupt reflective thinking, creative reasoning, and personal responsibility. Instead of replacing human cognition, AI was seen as reshaping mental demands and requiring new competencies in critical evaluation, meta-awareness, and calibrated reliance. The findings reveal the psychological dynamics of human-AI collaboration. They highlight the need for system designs and organisational practices that preserve human judgment and cognitive vitality. The study discusses implications for cognitive psychology, decision science, and applied professional practice.

Keywords

Human Agency, Epistemic Vigilance, Cognitive Offloading, Human-AI Collaboration, Qualitative Psychology, Decision-Making

1. Introduction

Artificial Intelligence (AI) is increasingly embedded in human decision-making

and cognitive activity, prompting extensive research across cognitive psychology, human-computer interaction, and organisational behaviour (Gonzalez & Heidari, 2025). While many studies focus on trust, accuracy, or usability, a growing body of work highlights the need to understand how AI reshapes human cognition, autonomy, and epistemic processes. (Valdez et al., 2024: p. 7). The speed and sophistication of contemporary AI systems evoke longstanding questions from cognitive psychology about agency, autonomy, attention, judgment, and the conditions under which humans maintain or relinquish control of their own thinking processes (Prunkl, 2022). AI's promise is well-documented, and research highlights that AI can significantly reduce cognitive load, accelerate information processing, and support users in synthesising large volumes of data (Buschmeyer et al., 2023; Liefooghe & van Maanen, 2023). In organisational or clinical settings, AI can condense complex reports, draft communications, and provide preliminary analyses with remarkable speed (Wang et al., 2024). For many professionals, these capabilities offer psychological relief by lowering task difficulty and enabling a focus on higher-order reasoning, emotional labour, interpersonal work, or creative problem solving (Dillon et al., 2025). Yet, alongside these benefits lies a set of less visible but psychologically consequential risks: diminished critical scrutiny, excessive cognitive offloading, dependencies that erode metacognitive skills, and overreliance on systems whose outputs may be flawed, biased, or unverifiable (Xiao et al., 2024).

As AI tools proliferate, psychologists have begun emphasising the need to examine not only trust or usability, but also human agency—the capacity for intentional, self-directed, reflective action within AI-supported environments (Bandura, 2001). Agency is not merely the freedom to act; it involves the psychological ownership of decisions, the ability to monitor and evaluate information critically, and the responsibility to correct or override external inputs (Satyanarayan & Jones, 2024). When humans rely on AI-generated outputs without adequate scrutiny, agency risks becoming diluted. Conversely, when humans remain vigilant—verifying information, examining sources, questioning anomalies—they preserve their cognitive and epistemic autonomy (Legare et al., 2018).

This leads to the second foundational psychological construct relevant to human-AI interaction: epistemic vigilance (Sperber et al., 2010). Epistemic vigilance refers to the innate cognitive mechanisms humans use to evaluate the truthfulness, reliability, and intent behind information provided by others (Sperber et al., 2010). Historically, these mechanisms evolved for social communication—to detect deception, persuasion, or error. Yet AI introduces a fundamentally different communication partner: one that simulates coherence, authority, and expertise without possessing genuine knowledge, accountability, or mental states (Weitz et al., 2021). Because AI outputs are often linguistically polished and delivered with apparent confidence, they can bypass intuitive scepticism and lead to unearned trust (Earp et al., 2025). As a result, epistemic vigilance becomes not just a cognitive disposition but a psychological necessity for safe and effective AI use (Singh et al.,

2025).

Within this theoretical backdrop, AI can be conceptualised not merely as a tool but as a cognitive environment that shapes how humans allocate attention, distribute mental effort, and assert or relinquish epistemic authority (Noller, 2025). Psychology is uniquely positioned to examine this environment, as it interrogates the internal mechanisms—attention, memory, metacognition, reasoning, and self-regulation—that underlie human interactions with external systems (Kahneman, 2011). Furthermore, psychological inquiry addresses the ethical dimension of AI, recognising that systems influencing cognition must be designed to support—not erode—reflective autonomy, judgment, and human dignity (Singh et al., 2025).

Cognitive boundaries refer to the perceived and enacted limits individuals construct between what they regard as their own cognitive responsibility and what can appropriately be delegated to an external system, such as AI. These boundaries are dynamic rather than fixed and are continuously negotiated in AI-supported environments where tasks traditionally requiring human reasoning, synthesis, or creative formulation can be offloaded (Collins et al., 2024). Closely related, recalibration of mental effort refers to the process through which individuals redistribute cognitive labour in response to AI's presence—often shifting effort away from generation and toward monitoring, verification, and evaluation (Everson, 2025).

These constructs informed both the interview prompts and the analytic lens. Participants were asked how they determined what should remain in their thinking when they felt the need to verify AI outputs, and how AI influenced their perceived cognitive effort, imagination, and sense of authorship. During analysis, particular attention was given to language reflecting boundary-setting, effort redistribution, scrutiny, and metacognitive regulation (Drosos et al., 2025).

Despite increased interest, substantial gaps remain. Many studies on AI in professional settings rely on quantitative surveys or experimental paradigms that measure trust, acceptance, accuracy, or decision outcomes but fail to capture the lived, subjective experience of negotiating agency and vigilance within real-world contexts (Papyshev, 2024: p. 2). Qualitative inquiry can address this gap by illuminating how professionals navigate uncertainty, adjust cognitive boundaries, engage in verification practices, and experience the psychological tensions between convenience and caution. Such an approach foregrounds human meaning-making rather than mechanical performance (Delacroix et al., 2025).

The present study draws upon rich qualitative data from professionals across psychology, leadership, and technology—individuals who engage daily with AI tools in settings where errors may have ethical, interpersonal, or organisational consequences (Everson, 2025). This dataset provides unique insight into how users consciously and unconsciously regulate their cognitive engagement with AI, how they preserve or surrender agency, and how they negotiate trustworthiness amid AI's fallibility (Shen et al., 2024: p. 17). Through a psychology-focused lens, this study investigates the inner negotiations that shape collaborative thinking between humans and machines (Chen et al., 2025).

The objective of this study was to examine how professionals in psychology, leadership, and technology psychologically experience, negotiate, and maintain human agency, epistemic vigilance, and cognitive boundaries when interacting with AI systems in cognitively demanding professional contexts. To explore how human agency, epistemic vigilance, and cognitive boundaries are psychologically experienced, negotiated, and maintained by professionals when interacting with AI.

2. Research Questions

* How do professionals perceive and preserve human agency—defined as the capacity for intentional action and decision-making—when using AI in cognitively demanding tasks?

* In what ways do users engage in epistemic vigilance—meaning the critical evaluation and monitoring of information sources—as well as verification and scrutiny when evaluating AI-generated outputs?

* How do individuals describe the shifting boundaries between their own cognitive effort and AI-enabled cognitive offloading, the latter referring to the process of delegating mental tasks to external systems?

These questions serve as a framework for exploring the psychological mechanisms underlying contemporary human-AI collaboration. They also set the stage for a review of relevant literature, where each research question is informed by findings in cognitive and applied psychology and their implications for theory, practice, and future investigations.

3. Literature Review

This literature review synthesises theories and empirical findings on human agency, epistemic vigilance, cognitive offloading, and mental boundary regulation, arguing that psychological constructs provide the foundation for understanding how individuals navigate AI-supported environments (Riley et al., 2025: p. 1).

3.1. Human Agency in Automated Cognitive Environments

Human agency is an individual's capacity to act intentionally, make choices, and make one's own decisions (Bandura, 2001). Traditional psychological models view agency as central to human functioning, supporting reflective judgment, volitional action, and responsibility for outcomes (Charisi & Dignum, 2024: p. 235). However, AI-mediated contexts prompt reconsideration. In these settings, the division of cognitive labor between humans and machines complicates agency (Vallor & Vierkant, 2024). AI's linguistic fluency and speed may lead users to accept outputs uncritically, raising questions about psychological ownership of decisions (Rahwan et al., 2019).

Rahwan et al. (2019) introduced the concept of society-in-the-loop, suggesting that human agency becomes distributed across technological systems, societal norms and algorithmic processes (Awad et al., 2020). In such environments, responsibility is often diffuse: humans rely on AI for cognitive scaffolding, while AI

depends on human guidance, prompting and oversight (Porter et al., 2024). This interdependence can support agency—when humans deliberately engage with AI as a collaborative cognitive tool—or undermine it when humans defer judgment due to system fluency or perceived expertise (Collins et al., 2024).

Research consistently demonstrates that humans attribute credibility and competence to AI systems, even when outputs are not verifiable (Peters & Visser, 2023). Delegating tasks to AI can initially support the agency by reducing cognitive strain; however, habitual delegation may weaken humans' sense of authorship and decision-making ownership (Alter & Oppenheimer, 2009). These dynamics underscore the importance of understanding how professionals navigate agency in AI-mediated contexts (Bauer et al., 2023: p. 5).

3.2. Epistemic Vigilance and the Psychology of Scrutiny

Epistemic vigilance was first conceptualised by Sperber et al. (Bender & Koller, 2020). Humans may misattribute intentionality or knowledge to AI—a cognitive bias known as anthropomorphic projection. (Kizilcec, 2023). When source transparency is absent, epistemic vigilance must compensate, requiring users to manually cross-check information. Xiao et al. (2024) showed that professionals experience significant cognitive and emotional disruption when encountering AI hallucinations or fabricated citations (Sun et al., 2024). This reinforces the need for psychological engagement through skepticism, cross-validation and domain expertise (Huemmer et al., 2025).

Conversely, seeing too many AI-generated insights can make people too confident in the information they read. Studies show that when AI is often right, people start to trust it too much, even for tasks it may not handle well (Glikson & Woolley, 2020). This shows the delicate balance between trusting and questioning AI when making decisions.

4. Methodological Approach

4.1. Research Design & Methodology

This study employed a qualitative research design grounded in reflexive thematic analysis (Braun & Clarke, 2006) to explore how professionals experience human agency, epistemic vigilance and cognitive boundaries in their interactions with AI systems. Qualitative inquiry was chosen because the study aimed to capture nuanced, subjective experiences that cannot be meaningfully quantified. A reflexive approach further supported the interpretive nature of this work, recognising the researcher as an active meaning-maker who engages with the data rather than functioning as a neutral observer. This design aligns with contemporary psychological research emphasising depth, context and cognitive-emotional processes in technology use (Brails, 2025).

4.2. Participants

Twelve participants were recruited through professional networks and purposive

sampling, ensuring diversity in cognitive demands and AI exposure. They represented three sectors where AI increasingly supports decision-making.

- Psychology professionals (n = 4)
- Technology professionals (n = 4)
- Leadership professionals (n = 4)

Participants ranged in age from 25 to 64 and all had at least 5 years of experience in their respective fields. They reported regular use of AI tools for tasks such as drafting reports, synthesising information, decision support, or cognitive scaffolding. Most participants were based in Singapore, with two located internationally (Canada and Thailand), offering multicultural contextual variation. Sample adequacy was determined in accordance with reflexive thematic analysis principles, which prioritise depth of meaning, conceptual richness, and interpretive coherence over numerical saturation. The twelve interviews provided sufficient cross-sector contrast to enable patterned comparisons across psychology, leadership, and technology domains. The dataset supported iterative theme refinement and theoretical development concerning agency preservation and cognitive boundary negotiation.

Participants reported regular engagement with generative AI tools such as ChatGPT, Microsoft Copilot, and domain-specific analytical systems. Use-cases differed across sectors: psychologists primarily used AI for report drafting and synthesis; leadership professionals for policy summarisation, strategic framing, and communication drafting; technology professionals for coding assistance, documentation, and troubleshooting. Frequency ranged from occasional task-based use to daily integration into workflow. These contextual variations shaped how vigilance, cognitive offloading, and boundary-setting were enacted (Araujo et al., 2020).

4.3. Inclusion & Exclusion Criteria

Inclusion criteria required professional experience with AI-enabled tools and English proficiency, as interviews were in English. Exclusion criteria eliminated those with no current AI exposure or who declined consent. To ensure confidentiality, pseudonyms and codes (e.g., AI5-Code 87) were used.

4.4. Procedure

A semi-structured interview schedule was developed to elicit participants' cognitive and psychological experiences when using AI. The schedule included open-ended prompts exploring:

- perceptions of agency and decision-making
- experiences of verification and scrutiny
- feelings of reliance or caution
- reflections on cognitive effort and imagination
- perceptions of AI's strengths and vulnerabilities

Academic supervisors reviewed the interview guide for clarity and alignment

with research aims. Interviews were conducted via Zoom, recorded with consent and transcribed verbatim. All interviews were conducted by the first author, who did not hold any supervisory or evaluative authority over participants. While some participants were recruited through extended professional networks, no direct reporting relationships existed. The interviewer maintained a reflexive stance throughout, acknowledging prior familiarity with AI systems and potential assumptions regarding cognitive risks or benefits. The semi-structured interview guide was piloted with two professionals outside the final sample to assess clarity, conceptual alignment, and flow. Minor revisions were made to simplify wording and ensure prompts elicited experiential narratives rather than abstract opinions.

Participants received an information sheet outlining the study's purpose, ethical rights, confidentiality measures and withdrawal procedures. Those who agreed to participate provided written informed consent electronically. Interviews were conducted individually and lasted 45 - 60 minutes.

A conversational yet structured approach allowed participants to expand on thoughts spontaneously while ensuring coverage of key research constructs. Interviews occurred within a one-week window to minimise variability in public discourse or technological change that might affect responses. Immediately after each interview, field notes captured reflexive impressions, emerging patterns and contextual factors relevant to interpretation. All transcripts were anonymised prior to analysis.

4.5. Data Analysis

Data were analysed using reflexive thematic analysis, following Braun and Clarke's (2006) six-phase structure:

* Familiarisation—reading transcripts multiple times and engaging with field notes.

* Initial coding—generating inductive and deductive codes related to agency, vigilance, cognitive effort and AI interactions.

* Theme development—grouping codes into meaningful clusters reflecting patterns across participants.

* Theme review—refining coherence within themes and distinctiveness between them.

* Theme definition—articulating conceptual boundaries and psychological meaning.

* Writing and integration—synthesising themes into a theoretical narrative.

Reflexivity was maintained throughout by acknowledging the researcher's prior experience with AI, potential biases and interpretive role. Rather than seeking inter-rater reliability, this reflexive approach emphasised transparency, coherence and theoretical depth—consistent with the epistemology of qualitative psychology (Heinrichs et al., 2025: p. 4). An audit trail was maintained throughout the analytic process, including detailed coding memos, reflexive annotations, and theme development diagrams documenting how initial codes evolved into higher-order

constructs. Interpretations were periodically discussed with academic supervisors to challenge assumptions, refine conceptual clarity, and enhance interpretive transparency. These discussions functioned as analytic debriefs consistent with reflexive thematic analysis rather than reliability testing (Buschmeyer et al., 2023).

4.6. Ethical Procedure

London Metropolitan University/Aventis granted ethical approval. Participants were informed of their rights to confidentiality, voluntary participation and withdrawal without consequence. Data were anonymised, securely stored and used only for research. Because of participants' professional roles, care was taken to ensure quotes could not identify individuals or organisations. The study followed the British Psychological Society's (2021) ethical guidelines.

4.7. Thematic Analysis

Reflexive thematic analysis generated two superordinate themes that capture how participants negotiate their cognitive and psychological relationship with AI. These themes are as follows: 1) Human Agency and Epistemic Vigilance and 2) Cognitive Boundaries and the Recalibration of Mental Effort.

Each theme contains multiple subthemes that illuminate how users maintain responsibility, exert scrutiny, regulate reliance and perceive the shifting landscape of human cognition in AI-augmented environments.

Theme 1: Human Agency and Epistemic Vigilance

Across all professional groups, participants emphasised that interacting with AI required maintaining psychological ownership of decisions. AI was described as useful but fundamentally fallible; this awareness triggered vigilant, active engagement rather than passive acceptance. Participants' narratives reflected continuous negotiation between trusting AI for efficiency and retaining agency through verification and oversight.

Subtheme 1.1: AI as a Starting Point, Not a Decision-Maker

Participants consistently rejected the idea that AI could replace human judgment. Instead, they framed AI as a "first draft," "assistive layer," or "starting point" that required human correction or refinement.

One psychologist described a moment in which she used AI to accelerate a clinical report but later discovered inaccuracies:

"Once I was in a hurry, I just used the AI output and added this to a report and then later on, I realised after submitting it, it's wrong… So then I realised, no, it's not a tool you can rely on fully when you need accuracy." (AI5-Code 87)

This experience became a catalyst for strengthened vigilance. AI's speed created a temptation to delegate cognitive responsibility, but errors reinforced the need for human oversight.

Similarly, a leadership professional emphasised:

"It gives you a head start, but you cannot let it lead the thinking. You still need to be the one deciding what makes sense." (AI3-Code 142)

This highlights a psychological stance wherein humans position themselves as intentional agents who must govern the cognitive contributions of AI.

Subtheme 1.2: Verification as Psychological Safeguard

Verification emerged as a central mechanism through which participants preserved agency. Regardless of profession, they described "double-checking," "triangulating," or "cross-verifying" AI outputs with traditional sources.

A leadership professional described consciously pairing AI with older methods:

"I'm still inclined to use AI but leverage my traditional data sources to double-check or fact-check what's come back." (AI4-Code 165)

A technology participant shared frustration after AI fabricated references:

"It may give something with a scientific-looking reference… and if you take that as valid without verifying, you are gone. I spent two hours looking for that paper… it sent me on a goose chase." (AI8-Code 126)

Verification thus acted as a *psychological safeguard*—a defence against cognitive complacency and a way to maintain intellectual ownership.

A contrasting perspective emerged from one technology participant who reported relatively high baseline trust in AI for routine coding tasks: "For boilerplate functions, I don't always double-check line by line—it's faster to test the output directly." (AI9-Code 54). However, this participant emphasised stricter scrutiny for high-stakes or client-facing deliverables. This divergent case illustrates that epistemic vigilance fluctuates depending on perceived task risk and domain familiarity (Collins et al., 2024).

Subtheme 1.3: Desire for Transparency and Traceable Sources

Participants repeatedly expressed that trustworthy AI must demonstrate where information came from. Transparency was linked to perceived reliability and cognitive confidence.

A psychologist noted:

"If the AI can give a reference link to where the information has come from, then the reliability improves." (AI5-Code 178)

Technology professionals were especially concerned:

"If there's one thing to change, it's making sure sources are validated. Fake citations cause the biggest problem." (AI11-Code 217)

Leadership participants echoed this need due to the consequences of presenting incorrect information:

"It needs to show legitimacy. Otherwise, how do I justify using it in important decisions?" (AI3-Code 214)

For participants, transparency reduced cognitive uncertainty and restored equilibrium between trust and vigilance.

Superordinate Theme 2: Cognitive Boundaries and the Recalibration of Mental Effort

The second superordinate theme captures how participants renegotiate the boundaries of their own cognitive labour in the presence of AI. They described both enhanced efficiency and emerging concerns about diminished imagination,

reduced persistence and overdependence.

Subtheme 2.1: Efficiency, Acceleration and Reduced Cognitive Strain

Participants universally acknowledged that AI dramatically reduced time and mental effort for information-heavy tasks.

A technology professional highlighted:

"I use the tool—probably I can just use five minutes to get the job done." (AI9-Code 33)

Another participant described how AI simplified policy interpretation:

"It reduced the time spent to understand policies and to get details needed." (AI12-Code 97)

Psychology participants found AI helpful for structuring thought:

"It gives me pointers and material to work with; I don't have to start from scratch." (AI2-Code 68)

Leaders emphasised emotional relief:

"You can write a thank you note through Copilot in seconds. That takes the mental load off." (AI4-Code 106)

These accounts demonstrate AI's strong capacity to reduce *extraneous cognitive load*—the effort associated with processing, summarising, or organising information.

Subtheme 2.2: The Emerging Risk of Cognitive Dulling

Alongside efficiency, participants voiced concerns about long-term impacts on creativity and critical thinking. They feared that habitual reliance on AI might dull internal cognitive capacities—especially imagination and spontaneous problem solving.

A psychologist noted:

"If you sit for too long using it, it affects your ability to think… your creative thinking becomes dull." (AI7-Code 63)

A technology professional provided a broader societal reflection:

"Use of the mind is slowly decreasing… human imagination is reducing with Gen AI." (AI11-Code 232)

Participants described a paradox: the very efficiency AI offers may reduce the need for humans to engage deeply, thereby diminishing cognitive sharpness. This mirrors psychological theories of "use-it-or-lose-it" cognitive maintenance and the risks associated with high-level cognitive offloading.

A leadership participant offered a nuanced counterpoint, noting that AI occasionally stimulated rather than dulled creativity: "Sometimes it gives angles I wouldn't have considered—it actually sparks ideas." (AI4-Code 118). This suggests that cognitive dulling is not inevitable but contingent on how AI is integrated—whether as replacement thinking or as generative provocation (Spatola, 2024).

Subtheme 2.3: Cognitive Boundaries Are Becoming Porous

Participants described shifting boundaries between what they "should" think through themselves and what AI could perform instead. Tasks previously requiring human effort—conceptualisation, phrasing, summarising—are now easily

delegated.

Yet participants also experienced discomfort when boundaries felt too porous:

"It's so easy to depend on it that sometimes I worry where to draw the line… what should still be my thinking." (AI10-Code 201)

Another stated:

"If you're not careful, it starts shaping the way you think, instead of the other way round." (AI6-Code 188)

This illustrates a meta-cognitive awareness: users recognise that AI influences their patterns of thought and must actively manage this influence to preserve autonomy.

Subtheme 2.4: Balancing Convenience with Cognitive Responsibility

Participants described the ongoing negotiation between convenience and responsibility. AI's usefulness was undeniable, but so were the ethical and psychological obligations to maintain cognitive effort.

A leadership professional reflected:

"It enhances your work, but you need to remain the primary thinker. AI should not become the default mind." (AI3-Code 140)

A psychologist framed it as a conscious discipline:

"I remind myself to think before looking at what AI proposes." (AI7-Code 74)

This negotiation aligns with psychological constructs of *self-regulation*, *meta-cognition* and *epistemic agency*—all essential for maintaining cognitive vitality in technology-rich environments.

5. Discussion

This study examined how professionals navigate agency, epistemic vigilance and cognitive boundaries when interacting with AI systems (Huemmer et al., 2025). The findings show that AI is not merely a tool; it serves as a cognitive environment that reshapes how individuals think, verify and manage mental effort (Georgiou, 2025). Two core dynamics emerged across participants: 1) active preservation of agency through verification and judgment and 2) recalibration of cognitive boundaries in response to AI-driven efficiency. These dynamics reveal deeper psychological processes underpinning modern human-AI interactions (Spatola, 2024).

The results showed that the agency is actively performing, not passively retained. Participants asserted authorship by verifying AI outputs, modifying suggestions, or treating AI as preliminary input rather than cognitive authority (Vodrahalli et al., 2021). This aligns with Bandura's (2001) view of agency as intentional, self-regulatory and reflective. Consistent with Rahwan et al. (2019), participants-maintained control by positioning themselves as final evaluators (Buder et al., 2021).

The strong emphasis on verification reflects a protective psychological strategy that preserves epistemic authority. Participants recognised that while AI offers accelerated cognitive processing, it also introduces risk through hallucinations, inaccuracies, or fabricated citations (Clark et al., 2025). These findings extend pre-

vious research showing that humans often overgeneralise trust in algorithmic fluency (Liao et al., 2020). In contrast, participants here displayed adaptive scepticism—a form of epistemic vigilance tailored to AI's unique combination of fluency and fallibility (Kadoma et al., 2024: p. 18).

Sperber et al.'s (2010) epistemic vigilance framework clarifies participants' behaviour. Because AI lacks comprehension but presents information confidently, users must rely on their own cognitive mechanisms to scrutinise claims. Participants described double-checking sources, cross-verifying content and spotting inconsistencies, making vigilance both a cognitive and emotional burden. AI reduced workload in some areas but increased it elsewhere, especially when users manually checked accuracy (Brails, 2025).

The psychological tension between efficiency and vigilance mirrors dual-process theories (Kahneman, 2011). Specifically, while AI encourages System 1 processing—fast, intuitive acceptance—its errors necessitate System 2 engagement—slow, analytical scrutiny. Managing this tension requires metacognitive awareness; accordingly, participants demonstrated this through intentional oversight of AI outputs (Hagendorff et al., 2023).

However, this vigilance was not without fatigue. Participants described frustration when AI errors undermine efficiency, reinforcing the need for responsible AI design that supports transparency, source attribution and error mitigation (Spatola, 2024).

AI's capacity for high-level cognitive offloading—drafting, structuring, summarising—shifted participants' views on which cognitive work requires human effort. While offloading theory (Risko & Gilbert, 2016) traditionally addresses memory or procedural tasks, here, participants offloaded conceptual and creative thinking. This broadens theoretical understanding of offloading into areas once seen as resistant to automation (Wahn & Schmitz, 2024: p. 2).

Participants valued AI's ability to reduce cognitive strain but simultaneously feared a decline in creativity, imagination and mental stamina (Gerlich, 2025). These concerns about "dulling" thought or becoming overly dependent suggest a need for clearer psychological boundaries in AI-supported cognition. Furthermore, these findings resonate with Buschmeyer et al. (2023), who warn that habitual use of AI may weaken cognitive resilience and reflective thinking over time (Georgiou, 2025).

The negotiation of boundaries—deciding what should remain human—emerged as both a psychological and ethical process. For example, users actively resisted overdependence by either thinking first or modifying AI outputs to ensure their own contribution (Drosos et al., 2025). This boundary maintenance, in turn, reflects a desire to protect individual identity as thinkers and decision-makers.

Participants' accounts indicate that epistemic vigilance is context-dependent rather than stable. Vigilance appeared most vulnerable under time pressure, routine task repetition, high prior confidence in AI outputs, and organisational expectations for rapid productivity. In such contexts, fast, intuitive acceptance of

AI-generated content may override deliberate verification (Gerlich, 2025). Participants described moments where deadlines reduced their scrutiny, particularly when AI had previously produced accurate results. These findings suggest that vigilance requires cognitive energy and situational awareness and may weaken when efficiency becomes prioritised over evaluation (Noller, 2025).

6. Applied Implications for Practice

The findings support several actionable implications. First, professional training programmes should explicitly cultivate verification literacy, including structured cross-check routines, source triangulation practices, and recognition of common hallucination patterns. Embedding verification competencies directly strengthens agency preservation (Theme 1). Second, organisations may implement workflow checkpoints for high-stakes outputs, requiring documented human validation before dissemination. Such safeguards externalise vigilance rather than relying solely on individual cognitive stamina.

Third, introducing deliberate friction strategies—such as prompting users to articulate their reasoning before consulting AI—may help preserve cognitive engagement and prevent boundary erosion (Theme 2). These interventions reinforce AI as an augmentative tool rather than a cognitive substitute (Wang et al., 2024).

7. Limitations and Directions for Future Research

While this study offers rich psychological insight into human agency, epistemic vigilance and cognitive boundary management in AI-mediated work, several limitations must be acknowledged. First, the sample size was relatively small (n = 12) and purposively selected. Although this aligns with qualitative methodological principles prioritising depth over breadth, the findings cannot be generalised statistically across all professional populations. Instead, they should be understood as theoretically transferable, offering conceptual insights that may resonate with similar high-stakes professional contexts.

Second, participants were predominantly experienced professionals who already demonstrated reflective awareness and critical engagement with AI. This may have biased the findings toward more vigilant and agentic patterns of AI use. Less experienced users, students, or individuals in routine or highly automated roles may exhibit different cognitive dynamics, including greater reliance, lower scrutiny, or reduced metacognitive awareness. Future research should therefore explore how agency and epistemic vigilance develop across different levels of expertise, career stages and educational backgrounds (Wahn & Schmitz, 2024: p. 2).

Third, this study relied on self-reported experiences gathered through interviews. While this approach effectively captures subjective meaning-making, it may not fully reflect unconscious cognitive processes or actual behavioural patterns during real-time AI interaction. Participants' accounts of vigilance and boundary-setting may differ from how they act under time pressure, fatigue, or

organisational constraints (Springer & Whittaker, 2020). Combining qualitative interviews with observational methods, task-based studies, or digital trace data could strengthen understanding of how reported intentions align with enacted behaviour (Collins et al., 2024).

Additionally, the rapid evolution of AI technologies represents a temporal limitation. The findings reflect experiences with contemporary generative AI tools at a particular moment in technological development (Araujo et al., 2020). As AI systems become more autonomous, multimodal and embedded within organisational infrastructures, the psychological demands placed on users may change. Longitudinal research is needed to examine how prolonged exposure to AI affects cognitive resilience, creativity, epistemic responsibility and professional identity over time (Wang et al., 2024).

Future research should also expand beyond individual cognition to examine organisational, cultural and ethical dimensions of human-AI collaboration. Investigating how workplace norms, leadership expectations, training practices and accountability structures shape epistemic vigilance would provide valuable applied insight. Finally, intervention-focused studies exploring how design features—such as transparency cues, source traceability, or deliberate friction—can support sustained human agency represent a promising and necessary direction for future psychological research (Everson, 2025).

8. Conclusion

This study contributes to the growing body of psychological research by examining how professionals experience AI in roles that require judgment, creativity and responsibility. Participants saw AI as beneficial for reducing workload, accelerating tasks and providing structural guidance. Yet they also recognised risks, including reduced critical thinking, weakened imagination, dependency, and the need for constant verification (Gerlich, 2025). AI's dual nature—efficient yet fallible—requires users to adopt an active cognitive stance characterised by scrutiny, responsibility and boundary-setting (Zhai et al., 2024: p. 27).

The study's core insight is that AI does not replace human cognition; rather, it reshapes it. Participants described becoming more vigilant, more intentional, and, depending on how AI was used, sometimes more cognitively passive. This negotiation between convenience and agency forms the psychological heart of AI-mediated cognition (Fernandes et al., 2024: p. 19).

Conflicts of Interest

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

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the Tribes of Fluency to Form a Metacognitive Nation. *Personality and Social Psychology Review, 13,* 219-235. https://doi.org/10.1177/1088868309341564

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI We Trust? Perceptions about Automated Decision-Making by Artificial Intelligence. *AI & Society, 35,* 611-623. https://doi.org/10.1007/s00146-019-00931-w

Awad, E., Dsouza, S., Bonnefon, J., Shariff, A., & Rahwan, I. (2020). Crowdsourcing Moral Machines. *Communications of the ACM, 63,* 48-55. https://doi.org/10.1145/3339904

Bandura, A. (2001). Social Cognitive Theory: An Agentic Perspective. *Annual Review of Psychology, 52,* 1-26. https://doi.org/10.1146/annurev.psych.52.1.1

Bauer, K., von Zahn, M., & Hinz, O. (2023). Please Take Over: Xai, Delegation of Authority, and Domain Knowledge. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4512594

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5185-5198). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.463

Brailas, A. (2025). Artificial Intelligence in Qualitative Research: Beyond Outsourcing Data Analysis to the Machine. *Psychology International, 7,* Article No. 78. https://doi.org/10.3390/psycholint7030078

Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology, 3,* 77-101. https://doi.org/10.1191/1478088706qp063oa

Buder, F., Pauwels, K., & Daikoku, K. (2021). The Illusion of Free Choice in the Age of Augmented Decisions. *NIM Marketing Intelligence Review, 13,* 46-51. https://doi.org/10.2478/nimmir-2021-0008

Buschmeyer, A., Krause, D., & Zander, M. (2023). Cognitive Offloading to AI Systems: Impacts on Critical Thinking and Problem Solving. *Journal of Cognitive Engineering and Decision Making, 17,* 95-110.

Charisi, V., & Dignum, V. (2024). Operationalizing AI Regulatory Sandboxes for Children's Rights and Wellbeing. In *Human-Centered AI* (pp. 231-249). Chapman and Hall/CRC. https://doi.org/10.1201/9781003320791-25

Chen, Z., Luo, Y., & Sra, M. (2025). *Engaging with AI: How Interface Design Shapes Human-AI Collaboration in High-Stakes Decision-Making*.

Clark, N., Shen, H., Howe, B., & Mitra, T. (2025). *Epistemic Alignment: A Mediating Framework for User-LLM Knowledge Delivery*.

Collins, K. M., Sucholutsky, I., Bhatt, U., Chandra, K., Wong, L., Lee, M. et al. (2024). Building Machines That Learn and Think with People. *Nature Human Behaviour, 8,* 1851-1863. https://doi.org/10.1038/s41562-024-01991-9

Delacroix, S., Robinson, D., Bhatt, U., Domenicucci, J., Montgomery, J., Varoquaux, G. et al. (2025). Beyond Quantification: Navigating Uncertainty in Professional AI Systems. *RSS: Data Science and Artificial Intelligence, 1,* udaf002. https://doi.org/10.1093/rssdat/udaf002

Dillon, E. W., Jaffe, S., Peng, S., & Cambon, A. (2025). *Early Impacts of M365 Copilot*.

Drosos, I., Sarkar, A., Xiaotong, & Toronto, N. (2025). *"It Makes You Think": Provocations Help Restore Critical Thinking to AI-Assisted Knowledge Work*.

Earp, B. D., Mann, S. P., Aboy, M., Clark, M. S. et al. (2025). *Relational Norms for Human-AI Cooperation*.

Everson, B. (2025). Exploring Employee Attitudes and Behaviors Related to AI Technology and the Future of Work. *Journal of Organizational Psychology, 25,* 6-18. https://doi.org/10.33423/jop.v25i2.7725

Fernandes, D., Villa, S., Nicholls, S., Haavisto, O., Buschek, D., Schmidt, A., Welsch, R. et

al. (2024). *Performance and Metacognition Disconnect When Reasoning in Human-AI Interaction*.

Georgiou, G. P. (2025). *ChatGPT Produces More "Lazy" Thinkers: Evidence of Cognitive Engagement Decline*.

Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies, 15,* Article No. 6. https://doi.org/10.3390/soc15010006

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals, 14,* 627-660. https://doi.org/10.5465/annals.2018.0057

Gonzalez, C., & Heidari, H. (2025). A Cognitive Approach to Human-AI Complementarity in Dynamic Decision-Making. *Nature Reviews Psychology, 4,* 808-822. https://doi.org/10.1038/s44159-025-00499-x

Hagendorff, T., Fabi, S., & Kosinski, M. (2023). Human-Like Intuitive Behavior and Reasoning Biases Emerged in Large Language Models but Disappeared in ChatGPT. *Nature Computational Science, 3,* 833-838. https://doi.org/10.1038/s43588-023-00527-x

Heinrichs, H., Kies, A., Nagel, S. K., & Kiessling, F. (2025). Physicians' Attitudes toward Artificial Intelligence in Medicine: Mixed Methods Survey and Interview Study. *Journal of Medical Internet Research, 27,* e74187. https://doi.org/10.2196/74187

Huemmer, M., Shyiramunda, T., Durner, F., & Cummings-Koether, M. J. (2025). *On the Influence of AI on Human Problem Solving*.

Kadoma, K., Metaxa, D., & Naaman, M. (2024). *Generative AI and Perceptual Harms*.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus, and Giroux.

Kizilcec, R. F. (2023). How Much Transparency Is Enough? *Computers in Human Behavior, 139,* Article ID: 107536.

Legare, C. H., Dale, M. T., Kim, S. Y., & Deák, G. O. (2018). Cultural Variation in Cognitive Flexibility Reveals Diversity in the Development of Executive Functions. *Scientific Reports, 8,* Article No. 16326. https://doi.org/10.1038/s41598-018-34756-2

Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-15). ACM. https://doi.org/10.1145/3313831.3376590

Liefooghe, B., & van Maanen, L. (2023). Cognitive Support through Automation. *Psychonomic Bulletin & Review, 30,* 234-252.

Noller, J. (2025). 4E Cognition and the Coevolution of Human-AI Interaction. *Discover Artificial Intelligence, 5,* Article No. 323. https://doi.org/10.1007/s44163-025-00595-0

Papyshev, G. (2024). *Governing AI through Interaction*. AI and Ethics.

Peters, T., & Visser, I. (2023). Human Skepticism and Overreliance. *Journal of Behavioral Decision Making, 36,* 54-70.

Porter, Z., Ryan, P., Morgan, P., Habli, I. et al. (2024). Unravelling Responsibility for AI. *SSRN*.

Prunkl, C. (2022). Human Autonomy in the Age of Artificial Intelligence. *Nature Machine Intelligence, 4,* 99-101. https://doi.org/10.1038/s42256-022-00449-9

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J., Breazeal, C. et al. (2019). Machine Behaviour. *Nature, 568,* 477-486. https://doi.org/10.1038/s41586-019-1138-y

Riley, C., Alrefai, O., Reyes, Y. C., & Hammad, E. (2025). *Human-AI Interactions*.

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

676-688. https://doi.org/10.1016/j.tics.2016.07.002

Satyanarayan, A., & Jones, G. M. (2024). *Intelligence as Agency*.

Shen, H., Knearem, T., Ghosh, R., Jurgens, D. et al. (2024). *Towards Bidirectional Human-AI Alignment*.

Singh, A., Taneja, K., Guan, Z., & Ghosh, A. (2025). *Protecting Human Cognition in the Age of AI*.

Spatola, N. (2024). The Efficiency-Accountability Tradeoff in AI Integration: Effects on Human Performance and Over-reliance. *Computers in Human Behavior: Artificial Humans, 2,* Article ID: 100099. https://doi.org/10.1016/j.chbah.2024.100099

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G. et al. (2010). Epistemic Vigilance. *Mind & Language, 25,* 359-393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Springer, S., & Whittaker, S. (2020). Mindful Delegation: Cognitive Boundary Management in Human-AI Collaboration. *International Journal of Human-Computer Studies, 142,* Article ID: 102463.

Sun, Y., Sheng, D., Zhou, Z., & Wu, Y. (2024). AI Hallucination: Towards a Comprehensive Classification of Distorted Information in Artificial Intelligence-Generated Content. *Humanities and Social Sciences Communications, 11,* Article No. 1278. https://doi.org/10.1057/s41599-024-03811-x

Valdez, A. C., Heine, M., Franke, T., Jochems, N., Jetter, H., & Schrills, T. (2024). The European Commitment to Human-Centered Technology: The Integral Role of HCI in the EU AI Act's Success. *i-com, 23,* 249-261. https://doi.org/10.1515/icom-2024-0014

Vallor, S., & Vierkant, T. (2024). Find the Gap: AI, Responsible Agency and Vulnerability. *Minds and Machines, 34,* Article No. 20. https://doi.org/10.1007/s11023-024-09674-0

Vodrahalli, K., Daneshjou, R., Gerstenberg, T., & Zou, J. (2021). Do Humans Trust Advice More If It Comes from AI? In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 763-777). ACM. https://doi.org/10.1145/3514094.3534150

Wahn, B., & Schmitz, L. (2024). A Bonus Task Boosts People's Willingness to Offload Cognition to an Algorithm. *Cognitive Research: Principles and Implications, 9,* Article No. 24. https://doi.org/10.1186/s41235-024-00550-0

Wang, Z., Cao, L., Danek, B., Jin, Q., Lu, Z., & Sun, J. (2024). Accelerating Clinical Evidence Synthesis with Large Language Models. *NPJ Digital Medicine, 8,* Article No. 509. https://doi.org/10.1038/s41746-025-01840-7

Weitz, K., Rosenberg, L., & Kraus, S. (2021). Anthropomorphic Biases in Human-AI Interaction. *Computers in Human Behavior, 124,* Article ID: 106935.

Xiao, Z., Liu, Y., & Fang, H. (2024). Emotional Consequences of AI Hallucinations. *Frontiers in Artificial Intelligence, 7,* 112-130.

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,* Article No. 28. https://doi.org/10.1186/s40561-024-00316-7