Advances in Social Sciences Research Journal – Vol. 12, No. 11
Publication Date: December 04, 2025
DOI:10.14738/assrj.1211.19642.
Khan, M. T., Fong, C., & Tripathi, S. (2025). "Trust, but Verify": A Reflexive Thematic Analysis of Human–AI Interaction. *Advances in Social Sciences Research Journal*, *12(11).* 237-251.

"Trust, but Verify": A Reflexive Thematic Analysis of Human–AI Interaction

Mohammad Tayyab Khan
School of Social Sciences,
London Metropolitan University, United Kingdom

Christopher Fong
School of Psychology,
University of Roehampton, United Kingdom

Shilpi Tripathi
Independent Researcher, Singapore

ABSTRACT

Artificial Intelligence (AI) has become deeply integrated into professional workflows, offering efficiency, scalability, and decision-support across sectors. Yet, questions remain about how users calibrate trust in AI and how reliance on these systems shapes human cognition. This study explores the psychological dimensions of trust, transparency, and cognitive load in human–AI interaction. Semi-structured interviews were conducted with twelve professionals across psychology, technology, and leadership domains. Data were analysed using Braun and Clarke's reflexive thematic analysis, revealing two superordinate themes: (1) trust as conditional, shaped by verification practices and expectations of source transparency, and (2) AI's dual role in reducing cognitive load while raising concerns about diminishing creativity and imagination. Findings highlight that professionals value AI as a supportive assistant that saves time and streamlines tasks but remain cautious about accuracy, hallucinations, and overreliance. The study contributes to qualitative research on human–AI interaction by emphasising the need for explainability, verifiable outputs, and safeguards against cognitive complacency. It recommends psychologically informed design strategies that balance efficiency with transparency and preserve users' epistemic agency.

Keywords: human–AI interaction, trust, explainable AI, cognitive load, cognitive offloading.

INTRODUCTION

Background

Artificial Intelligence (AI) has rapidly transitioned from being a niche technological innovation to a ubiquitous presence in modern professional, educational, and social life. AI no longer confined to the research laboratories of computer scientists or the experimental prototypes of technology companies, AI now permeates everyday practices across industries and sectors (Raees et al., 2024). Applications span automation in logistics, predictive analytics in finance, diagnostic assistance in healthcare, adaptive learning platforms in education, and decision-support tools in leadership and organizational management (Soomro et al., 2025). This broad integration reflects AI's dual promise: enhanced efficiency through automation and the potential to augment human cognitive processes by offering insights, recommendations, or new ways of engaging with data. However, alongside these technical advances lies an increasingly recognized reality, that the success of AI adoption depends as much on *psychological* and *behavioural* dimensions as on technical performance (Freitas et al., 2023).

Trust has emerged as a decisive determinant of whether and how humans adopt AI technologies. Trust influences not only initial acceptance but also long-term reliance and sustained engagement (Li et al., 2024). Liao et al. (2020) demonstrated that when AI outputs are transparent and verifiable, users are significantly more likely to incorporate AI-generated suggestions into their workflows. Conversely, when systems produce what are often termed "hallucinations", outputs that are grammatically fluent but factually inaccurate, users' confidence declines, producing two distinct but equally problematic behaviours: disengagement, where users reject AI altogether, or blind reliance, where users accept AI outputs without adequate scrutiny (Peters & Visser, 2023). Both extremes are problematic, highlighting the delicate psychological balance required in cultivating appropriate levels of trust (Srinivasan & Thomason, 2025).

Johnson and Grayson (2005) provide a useful framework for understanding this balance, conceptualising trust as comprising two dimensions: *cognitive trust* and *affective trust*. Cognitive trust is grounded in perceptions of competence, reliability, and logical coherence. Users must believe that an AI system is accurate and consistent. Affective trust, by contrast, emerges from feelings of security, comfort, and relational warmth. In human–AI interactions, this may manifest when AI interfaces appear empathetic, user-friendly, or aligned with social expectations of helpfulness (McKee et al., 2023). For AI systems to be both usable and sustainable in professional contexts, designers must address both forms of trust. This involves building mechanisms for verifiability and accuracy, while also fostering positive emotional experiences that encourage sustained engagement (Xu et al., 2024).

Transparency is a cornerstone of building trust. A common criticism of AI, particularly of large-scale machine learning models, is their "black-box" nature. Users often cannot see how an output was generated or which data points influenced a given prediction (Ramachandram et al., 2025). Explainable AI (XAI) seeks to address this by providing rationales, confidence scores, or references that clarify the reasoning process behind an output (Liao et al., 2020). In professional contexts, this transparency is not merely an add-on but a prerequisite. For instance, in healthcare, clinicians must understand and trust the reasoning behind an AI-assisted diagnosis before integrating it into patient care. Similarly, in leadership contexts, executives require verifiable justifications for decisions, particularly when accountability and reputation are at stake. The absence of transparency undermines trust and may limit AI's broader adoption (Kaufman & Kirsh, 2023).

Another crucial psychological dimension shaping human–AI interaction is cognitive load. Cognitive Load Theory (Sweller, 1988) provides a framework for understanding how humans process information and allocate mental resources. The theory differentiates between three types of load: intrinsic, extraneous, and germane. *Intrinsic load* relates to the inherent complexity of the task; *extraneous load* arises from the way information is presented or the usability of the interface; and *germane load* reflects the cognitive resources devoted to meaningful learning or problem-solving (Merriënboer & Sweller, 2009).

AI systems have the potential to reduce extraneous cognitive load by automating repetitive, data-heavy, or administrative tasks. For example, AI-driven summarisation tools can quickly condense long documents, allowing professionals to focus on interpretation rather than information gathering (Watermeyer et al., 2023). Decision-support systems can filter through vast datasets, highlighting relevant patterns or anomalies that would otherwise overwhelm human cognition. In this sense, AI can act as a cognitive partner, freeing mental resources for higher-order reasoning, creativity, or strategic decision-making (Liefooghe & van Maanen, 2023).

However, this very strength introduces a paradox. By consistently outsourcing routine or even moderately complex tasks to AI, users may experience what scholars term *cognitive offloading,* a reliance on external systems that diminishes the development or exercise of internal cognitive skills (Gerlich, 2025). Buschmeyer et al. (2023) caution that while AI reduces immediate mental strain, long-term dependence may erode critical thinking, problem-solving, and imaginative capacity. Professionals in education and psychology echo this concern, warning that overreliance could lead to a generation of workers less adept at reflective reasoning or creative ideation. In leadership contexts, the risk is even more pronounced: decision-makers who rely heavily on AI recommendations may lose their ability to critically evaluate alternatives, compromising both autonomy and accountability (Gerlich, 2025).

The challenge, then, lies in balancing AI's efficiency with the preservation of human cognitive engagement. Sperber et al.'s (2010) concept of *epistemic vigilance* is instructive here (Lai et al., 2025). Humans naturally scrutinise information sources to guard against deception or error. In the context of AI, this vigilance must be deliberately nurtured rather than eroded. Users should remain active participants in the interpretive process, verifying outputs, questioning assumptions, and applying professional judgment (Bignami et al., 2025). Without this balance, the very efficiency gains offered by AI may inadvertently foster complacency, reducing critical oversight and potentially leading to errors with significant consequences (Al-Zahrani, 2024).

It is also important to recognise that psychological responses to AI vary across professional domains. Psychologists may focus on the ethical and relational dimensions, emphasising safeguards against manipulation or bias (Naseer et al., 2025). Technology professionals may prioritise functional accuracy, demanding robust performance and reliability. Leaders, on the other hand, may emphasise accountability and legitimacy, particularly given the reputational risks associated with decision-making (Nguyen & Shaik, 2024). These variations suggest that AI design cannot adopt a one-size-fits-all approach. Instead, it must be context-sensitive, accommodating the specific expectations, risks, and psychological orientations of different professional groups (Buschmeyer et al., 2023).

CURRENT GAPS IN EXISTING LITERATURE

Much of the existing literature on trust and cognition in AI remains quantitative, focusing on experimental tasks, accuracy rates, or user satisfaction surveys. While informative, these methods often overlook the nuanced, lived experiences of professionals who integrate AI into their daily practices (Bossen & Pine, 2022). Qualitative research offers a complementary

perspective, capturing the subtleties of conditional trust, verification practices, and ambivalence about cognitive offloading. By foregrounding the voices of users, qualitative inquiry can reveal how psychological theories manifest in real-world contexts, offering richer insights into the challenges and opportunities of human–AI collaboration (Jiang et al., 2021).
AI's future success is not determined solely by technical improvements but by how well systems account for human psychological dynamics. Trust, both cognitive and affective, must be cultivated through transparency and verification (Raees et al., 2024). Cognitive load must be carefully managed, ensuring that efficiency gains do not come at the expense of creativity and critical thinking. Professional contexts must be acknowledged as shaping diverse expectations of AI (Söllner et al., 2025). Acknowledging these factors is essential for advancing Human-Centred Artificial Intelligence (HCAI), where systems are designed not merely to automate tasks but to enhance human cognition, preserve autonomy, and foster sustainable engagement (Schmager et al., 2025).

Research Aims

This study addresses these gaps by examining how professionals across psychology, technology, and leadership sectors experience AI in relation to:

1. Conditional trust and verification practices.
2. Expectations of transparency and source reliability.
3. Perceived efficiency gains and reduced cognitive load.
4. Concerns about diminishing creativity and imagination.

By focusing on the psychological dimensions of trust and cognition, the study contributes to the design of human-centred AI systems that enhance usability while preserving epistemic vigilance and critical engagement (Kereopa-Yorke, 2025).

METHODOLOGY

Research Design

A qualitative research design was adopted to capture the nuanced and context-specific experiences of professionals engaging with AI in their daily work. Given that the study aimed to explore psychological dimensions such as trust, transparency, and cognitive effort, a qualitative approach was considered most appropriate, as it allows for depth, flexibility, and the inclusion of diverse perspectives. Semi-structured interviews were used as the primary method of data collection. This format enabled participants to describe their lived experiences in detail while allowing the researcher to probe for clarification and elaboration when necessary. Reflexive thematic analysis, as outlined by Braun and Clarke (2006), was chosen as the analytic framework. This method is particularly suited to qualitative inquiry because it accommodates both inductive theme generation from the data and deductive alignment with the study's conceptual focus on trust and cognition.

Participants

The study involved twelve participants drawn from three professional groups: psychology (n = 4), technology (n = 4), and leadership (n = 4). These groups were purposely selected to capture perspectives from fields where AI is integrated into decision-making, practice, and professional judgement. Each participant had at least five years of relevant professional experience and direct, ongoing engagement with AI tools, ensuring that their contributions reflected informed and sustained interaction with such systems. Recruitment was conducted via professional networks and snowball sampling, enabling access to experienced practitioners. Diversity across gender, role seniority, and sectoral context was sought to ensure a balanced representation of views. While most participants were based in Singapore, two were located internationally (Canada and Thailand), broadening the study's cultural and contextual scope.

Identifier	Gender	Age Range (Yrs)	Working Experience Range (Yrs)	Industry
AI1	M	45-54	More than 20	Leadership
AI2	F	45-54	More than 20	Leadership
AI3	F	45-54	More than 20	Leadership
AI4	M	45-54	More than 20	Leadership
AI5	F	45-54	More than 20	Psychology
AI6	M	55-64	More than 20	Psychology
AI7	F	25-34	6-10	Psychology
AI8	M	35-44	16-20	Psychology
AI9	F	35-44	More than 20	Technology
AI10	F	25-34	6-10	Technology
AI11	M	35-44	16-20	Technology
AI12	M	45-54	More than 20	Technology

Data Collection

Data were collected through online interviews conducted via Zoom, each lasting between 45 and 60 minutes. This mode of delivery was selected for accessibility and convenience, particularly given the global dispersion of participants. The interview guide included open-ended questions focusing on AI usability, perceptions of trust and transparency, efficiency, and cognitive impact. Probing questions encouraged participants to provide concrete examples and reflections on their professional use of AI. All interviews were audio-recorded with prior consent, transcribed verbatim, and anonymised. To ensure confidentiality, participant identities were replaced with unique identifiers (e.g., AI5 – Code 87). Field notes were taken after each interview to capture initial impressions and contextual factors (Townsend et al., 2023).

Data Analysis

Thematic analysis followed Braun and Clarke's (2006) six phases: (1) familiarisation with data, (2) generation of initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the final report. Analysis was iterative, with early coding conducted independently by the primary researcher before discussion with a peer reviewer to ensure reflexivity and reduce bias. Both inductive and deductive coding strategies were employed: inductive to capture unanticipated participant perspectives, and deductive to ensure alignment with the study's central concepts of trust, transparency, and cognitive load. A thematic matrix was created to map connections between codes, subthemes, and overarching themes. Verbatim quotes were selected to illustrate key points, ensuring participants' voices remained central to the analysis.

Ethical Considerations

The study received ethical approval from London Metropolitan University/Aventis. Ethical safeguards were rigorously applied in line with the British Psychological Society's (BPS, 2021) ethical guidelines. Participants were fully informed about the study's aims, procedures, and potential risks before providing written consent. They were reminded of their right to withdraw at any stage without penalty. To protect confidentiality, all identifying details were removed during transcription, anonymised codes were used in reporting, and data were securely stored on encrypted devices. Audio recordings were deleted after transcription, and only the anonymised text was retained for analysis.

FINDINGS

Two superordinate themes were identified: Trust and Reliability and Cognitive Load and Efficiency. Each contained subthemes that reflected both the perceived benefits and risks of AI in professional contexts.

Theme 1: Trust and Reliability

Subtheme 1.1: Conditional Trust and Verification:

Participants consistently described trust in AI as conditional, never absolute. They valued AI's efficiency but stressed that outputs required verification against credible sources (Fine & Marsh, 2024). One participant recounted a professional error that arose from misplaced reliance:

> *"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 which I can really rely on in terms of when you want that accuracy." (AI5 – Code 87)*

This illustrates a pragmatic stance, AI is seen as a starting point rather than a final authority. Another participant shared frustration with fabricated references, which consumed valuable time and undermined confidence (Xiao et al., 2024).

> *"..it may give you a certain output and a reference from what looks a like scientific paper and if you take that as valid without verifying, you are gone…I spent two hours looking for that paper on the internet…it sent me on a goose chase." (AI8 – Code 126)*

Others described how AI sometimes produced tangential or irrelevant responses, reinforcing their reliance on cross-checking:

> *"When we start asking it questions, it can just go off tangent and tell you something that you know is not related, or you know, is not even existent." (AI11 – Code 129)*

To manage this, many adopted hybrid practices, balancing AI's speed with human verification:

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

Taken together, these accounts reflect what Sperber et al. (2010) call epistemic vigilance, the human tendency to critically assess new information before acceptance. AI was valued as an assistant, but professionals resisted outsourcing judgement entirely (Bossen & Pine, 2022).

Subtheme 1.2: Citation and Source Transparency:

Trust was closely tied to transparency. Participants emphasised that AI outputs must be traceable to reliable sources. As one psychologist explained:

> *"If the AI is giving a specific output, if it can give a reference link to where the information has come from, then the reliability factor improves." (AI5 – Code 178)*

Technology professionals pointed to fictitious citations as the most damaging problem:

> *"If there's one thing we can change is for it to be able to validate the source, because the biggest problem is when it starts giving inaccurate information because of citing fictitious website." (AI11 – Code 217)*

Leaders especially stressed legitimacy, noting that unverified information could damage reputations:

> *"…references the source of that data. And I think that, for me, is important, because that gives a little bit more comfort and assurance that what is being provided is coming from somewhere legit and not complete fiction." (AI3 – Code 214)*

These insights align with explainable AI (XAI) research (Liao et al., 2020), which demonstrates that verifiable sources foster trust. For participants, transparency was not a "bonus feature" but a prerequisite for serious adoption.

Theme 2: Cognitive Load and Efficiency

Subtheme 2.1: Productivity and Efficiency Gains:

Participants spoke positively about AI's ability to reduce workload and save time. A common view was that AI streamlined tasks and simplified complex documents:

> *"I use the tool, probably I can just use five minutes to get the job done." (AI9 – Code 33)*
>
> *"Reduced the time spent to understand policies and to get details needed." (AI12 – Code 97)*
>
> *"it does help me, because I don't have to spend too much time in thinking…it's very efficient." (AI7 – Code 69)*

Many framed AI as providing useful starting points:

> *"Once the pointers are in… it gives me more material and content to work with." (AI2 – Code 68)*

Leaders noted its value for quick communications:

> *"…write a thank you note, you can pretty much write that through your Copilot and it's done in seconds. So that really frees up a lot of time." (AI4 – Code 106)*

Similarly, technologists praised reporting efficiency:

URL: http://dx.doi.org/10.14738/assrj.1211.19642

> *"AI can be very good at reporting… now within half an hour, everything is done." (AI8 – Code 117).Even when prompts were weak, AI was still viewed as reducing effort:*
>
> *"the way I was prompting the model to ask a question, I think that was also not very well structured as a result I had very little confidence on output." (AI12 – Code 135)*

The consensus was clear: AI lowered extraneous cognitive load (Sweller, 1988), freeing participants to focus on strategic, higher-level thinking. For some, the appeal was psychological as well as practical:

> *"Very easy... these tools are very addictive because they take the onerous aspect of a job out." (AI2 – Code 56)*
>
> *"It enhances it... gives you space to think ahead... more information to make more effective decisions." (AI4 – Code 100)*

Subtheme 2.2: Diminishing Human Cognition and Imagination:

Despite enthusiasm for efficiency, participants expressed caution about overreliance on AI. A psychologist warned:

> *"If you sit for a very long time, I think it could affect your ability to think and your creative thinking… it becomes dull." (AI7 – Code 63)*

A technologist echoed this, noting generational risks:

> *"The use of the mind is slowly decreasing… human imagination is reducing with Gen AI." (AI11 – Code 232)*

These concerns reflect the literature on cognitive offloading (Buschmeyer et al., 2023), which warns of dependency risks. While AI reduced workload in the present, participants worried it might erode critical and imaginative capacities in the long term.

Subtheme 1.2: Citation and Source Transparency:

A recurring concern was the lack of transparency in AI outputs. Professionals consistently framed citation and source referencing as critical to reliability (Spennemann, 2025). Without traceable references, AI was perceived as unreliable, even deceptive.

One psychologist explained:

> *"If the AI is giving a specific output, if it can give a reference link to where the information has come from, then the reliability factor improves." (AI5 – Code 178)*

For technology professionals, fabricated citations were viewed as the "biggest problem" because they undermined efficiency rather than enhancing it:

> *"If there's one thing we can change is for it to be able to validate the source, because the biggest problem is when it starts giving inaccurate information because of citing fictitious website." (AI11 – Code 217)*

Leaders echoed this sentiment, connecting source transparency with legitimacy and professional accountability:

> *"…references the source of that data. And I think that, for me, is important, because that gives a little bit more comfort and assurance that what is being provided is coming from somewhere legit and not complete fiction." (AI3 – Code 214)*

These reflections echo the Explainable AI (XAI) literature (Liao et al., 2020), which stresses that transparency mechanisms build trust by allowing users to interrogate and verify AI outputs. For participants, transparency was not an optional add-on but a precondition for integration into professional workflows.

Theme 2: Cognitive Load and Efficiency

Subtheme 2.1: Productivity and Efficiency Gains:

Participants widely agreed that AI reduced extraneous cognitive load by simplifying complex tasks, accelerating drafting, and cutting down repetitive work.

> *"I use the tool, probably I can just use five minutes to get the job done." (AI9 – Code 33)*
>
> *"Reduced the time spent to understand policies and to get details needed." (AI12 – Code 97)*
>
> *"it does help me, because I don't have to spend too much time in thinking…it's very efficient." (AI7 – Code 69)*

Many described AI as a springboard—providing "pointers" or drafts that could be refined further:

> *"Once the pointers are in… it gives me more material and content to work with." (AI2 – Code 68)*

Leaders valued its time-saving role in communications:

> *"…write a thank you note, you can pretty much write that through your Copilot and it's done in seconds. So that really frees up a lot of time." (AI4 – Code 106)*

Similarly, technologists celebrated its speed in reporting:

> *"AI can be very good at reporting… now within half an hour, everything is done." (AI8 – Code 117)*
>
> *Even when prompting was imperfect, AI was still perceived as a net efficiency gain:*
>
> *"the way I was prompting the model to ask a question, I think that was also not very well structured as a result I had very little confidence on output." (AI12 – Code 135)*

Participants repeatedly described AI as an enabler of higher-level thinking, echoing Cognitive Load Theory (Sweller, 1988):

> *"Very easy... these tools are very addictive because they take the onerous aspect of a job out." (AI2 – Code 56)*
>
> *"It enhances it... gives you space to think ahead... more information to make more effective decisions." (AI4 – Code 100)*

This demonstrates that AI was not only saving time but also reshaping how professionals allocated mental effort across tasks.

Subtheme 2.2: Diminishing Human Cognition and Imagination:

Despite their praise for efficiency, participants worried about overreliance on AI diminishing human creativity and reflective thought.

A psychologist expressed this concern clearly:

> *"If you sit for a very long time, I think it could affect your ability to think and your creative thinking… it becomes dull." (AI7 – Code 63)*

A technology professional echoed this sentiment, linking it to generational changes in thinking habits:

> *"The use of the mind is slowly decreasing… human imagination is reducing with Gen AI." (AI11 – Code 232)*

These reflections align with the literature on cognitive offloading (Buschmeyer et al., 2023), where repeated delegation of cognitive effort to external systems risks eroding long-term capacity for critical thinking and imagination. For participants, this represented a paradox: while AI amplified productivity in the short term, it risked dulling human ingenuity over time. The tension between efficiency and creativity illustrates the core dilemma of human–AI collaboration, AI as both an enabler and a potential inhibitor of cognitive development (Riva, 2025).

DISCUSSION

This study sheds light on the complex ways professionals engage with AI, particularly around the psychological dimensions of trust and cognition (Jose et al., 2025). The findings confirm that trust in AI is not static but conditional, calibrated through continuous verification, transparency of sources, and task context. This echoes the Affect–Cognition Trust Model (Johnson & Grayson, 2005), which highlights the interplay between rational assessments of reliability (cognitive trust) and emotional comfort or security (affective trust). Our data demonstrate that professionals do not accept AI outputs at face value; instead, they adopt hybrid practices where AI is used as a supportive tool but cross-checked against established sources. This reflects epistemic vigilance (Sperber et al., 2010), the tendency of humans to scrutinize new information for accuracy and legitimacy.

Importantly, trust was not homogeneous across professional domains. Leaders placed emphasis on accountability and reputational risks, psychologists foregrounded ethical safeguards, while technology professionals prioritized technical accuracy (Mahajan, 2025). These role-specific priorities suggest that trust in AI cannot be engineered through universal design features alone but must be tailored to the expectations and risks of specific user groups (Li et al., 2024).

The findings also underscore the dual role of AI in managing cognitive load. Participants valued AI's ability to reduce extraneous load (Sweller, 1988) by automating repetitive tasks, thereby freeing mental resources for higher-order reasoning. This aligns with prior research showing that AI can enhance productivity and efficiency by acting as a cognitive partner (Liefooghe & van Maanen, 2023). However, participants also expressed concern about cognitive offloading (Buschmeyer et al., 2023). Overreliance on AI was perceived as a threat to imagination, creativity, and critical thinking, raising the paradox that tools designed to support cognition may, over time, erode the very skills they seek to enhance. This ambivalence highlights the need to conceptualize AI as a scaffold rather than a substitute for human cognition (Jose et al., 2025).

Finally, the demand for explainability and source transparency was a recurrent theme. Participants consistently requested verifiable references and clear justifications for AI outputs, reflecting the importance of Explainable AI (XAI) principles (Liao et al., 2020). Concerns about hallucinations, or fabricated outputs presented with authority, underscore the fragility of trust when systems appear convincing but deliver falsehoods (Kerasidou, 2020). This reinforces that transparency is not optional but foundational for responsible AI deployment.

IMPLICATIONS

For AI Design

- Source verification: Embedding mechanisms that provide traceable references can transform trust from tentative to durable.
- Progressive disclosure: Interfaces should allow users to choose the level of explanation detail, enabling efficiency without overwhelming them.
- Safeguards against overreliance: Built-in nudges (e.g., prompts encouraging cross-checking) could mitigate cognitive complacency.

For Professional Practice

- Critical AI literacy: Training should emphasize not only how to use AI but also how to interrogate its outputs critically.
- Hybrid workflows: Encouraging practices that combine AI efficiency with human oversight can maintain accountability and preserve epistemic agency.

For Research

- Longitudinal studies: Future research should examine how extended AI use affects cognitive resilience, imagination, and trust calibration over time.
- Role-specific frameworks: Exploring how different professional groups (e.g., healthcare, law, education) prioritise trust and transparency can inform more targeted design of HCAI systems.

URL: http://dx.doi.org/10.14738/assrj.1211.19642

Cross-cultural perspectives: Since trust norms vary globally, comparative studies could deepen understanding of contextual differences in human–AI interaction.

CONCLUSION

This study demonstrates that professionals experience AI as a paradoxical presence, simultaneously a cognitive enabler and a potential cognitive eroder. Trust is not absolute but conditional, shaped by verification practices, source transparency, and professional context. Similarly, while AI reduces extraneous cognitive load and enhances productivity, it raises concerns about long-term dependence, diminished creativity, and the erosion of critical thinking (Gerlich, 2025).

The findings reinforce the importance of psychologically informed AI design, where efficiency is balanced with transparency, and user agency is preserved. By embedding verifiable sources, fostering explainability, and designing safeguards against overreliance, AI can function as a scaffold that enhances rather than diminishes human cognition (Singh et al., 2025).

Future research must extend these insights beyond immediate professional contexts, exploring diverse cultural and occupational settings and tracking long-term cognitive effects. Ultimately, the path forward lies in developing Human-Centred AI systems that not only improve efficiency but also sustain creativity, accountability, and trustworthiness, qualities essential for AI's responsible integration into professional and societal life.

References

Alabed, A., Javornik, A., & Gregory-Smith, D. (2022). AI anthropomorphism and its effect on users' self-congruence and self–AI integration: A theoretical framework and research agenda. *Technological Forecasting and Social Change, 182*, 121786. https://doi.org/10.1016/j.techfore.2022.121786

Al-Zahrani, A. M. (2024). Balancing act: Exploring the interplay between human judgment and artificial intelligence in problem-solving, creativity, and decision-making. *IgMin Research, 2*(3), 145. https://doi.org/10.61927/igmin158

Bara, C. D., Wang, X., & Chai, J. Y. (2021). MindCraft: Theory of mind modeling for situated dialogue in collaborative tasks. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 4857–4868). https://doi.org/10.18653/v1/2021.emnlp-main.396

Becker, D., Rueda, D., Beese, F., Torres, B. S. G., Lafdili, M., Ahrens, K., Fu, D., Strahl, E., Weber, T., & Wermter, S. (2023). The emotional dilemma: Influence of a human-like robot on trust and cooperation. *IEEE RO-MAN 2023 Conference Proceedings*, 1689–1696. https://doi.org/10.48550/arXiv.2307.02924

Bignami, E. G., Russo, M., Semeraro, F., & Bellini, V. (2025). Technoethics in real life: AI as a core clinical competency. *Journal of Anesthesia Analgesia and Critical Care, 5*(1). https://doi.org/10.1186/s44158-025-00233-2

Bornet, P., Wirtz, J., Davenport, T. H., De Cremer, D., Evergreen, B., Fersht, P., Gohel, R., Khiyara, S., Sund, P., & Mullakara, N. (2025). *Agentic artificial intelligence: Harnessing AI agents to reinvent business, work and life*. Irreplaceable Publishing.

Bossen, C., & Pine, K. H. (2022). Batman and Robin in healthcare knowledge work: Human–AI collaboration by clinical documentation integrity specialists. *ACM Transactions on Computer-Human Interaction, 30*(2), 1. https://doi.org/10.1145/3569892

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

British Psychological Society. (2021). *Code of ethics and conduct*. https://www.bps.org.uk/guideline/code-ethics-and-conduct

Buschmeyer, K., Hatfield, S., & Zenner, J. (2023). Psychological assessment of AI-based decision support systems: Tool development and expected benefits. *Frontiers in Artificial Intelligence, 6*, 1092053. https://doi.org/10.3389/frai.2022.1092053

De Filippis, S., & Al Foysal, M. J. (2024). Ethical vigilance in AI deployment. *AI & Society*. https://doi.org/10.1007/s00146-024-01687-2

European Commission. (2024). *Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. https://ec.europa.eu

Fine, A., & Marsh, S. C. (2024). Judicial leadership matters (yet again): The association between judge and public trust for artificial intelligence in courts. *Discover Artificial Intelligence, 4*(1). https://doi.org/10.1007/s44163-024-00142-3

Freitas, J. D., Agarwal, S., Schmitt, B. H., & Haslam, N. (2023). Psychological factors underlying attitudes toward AI tools. *Nature Human Behaviour, 7*(11), 1845. https://doi.org/10.1038/s41562-023-01734-2

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), 6. https://doi.org/10.3390/soc15010006

Jiang, J. A., Wade, K., Fiesler, C., & Brubaker, J. R. (2021). Supporting serendipity: Opportunities and challenges for human–AI collaboration in qualitative analysis. *arXiv*. http://arxiv.org/abs/2102.03702

Johnson, D., & Grayson, K. (2005). Cognitive and affective trust in service relationships. *Journal of Business Research, 58*(4), 500–507. https://doi.org/10.1016/S0148-2963(03)00140-1

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

Kaufman, R., & Kirsh, D. (2023). Explainable AI and visual reasoning: Insights from radiology. *arXiv*. https://doi.org/10.48550/arXiv.2304.03318

Kerasidou, A. (2020). Artificial intelligence and the ongoing need for empathy, compassion, and trust in healthcare. *Bulletin of the World Health Organization, 98*(4), 245–250. https://doi.org/10.2471/BLT.19.237198

Kereopa-Yorke, B. (2025). Engineering trust, creating vulnerability: A socio-technical analysis of AI interface design. *arXiv*. https://doi.org/10.48550/arXiv.2507.02866

Lai, S., Kim, J., Kunievsky, N., Potter, Y., & Evans, J. (2025). Biased AI improves human decision-making but reduces trust. *arXiv*. https://doi.org/10.48550/arXiv.2508.09297

Li, Y., Wu, B., Huang, Y., & Luan, S. (2024). Developing trustworthy artificial intelligence: Insights from research on interpersonal, human-automation, and human–AI trust. *Frontiers in Psychology, 15*. https://doi.org/10.3389/fpsyg.2024.1382693

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). https://doi.org/10.1145/3313831.3376590

Liefooghe, B., & van Maanen, L. (2023). Reducing cognitive load through adaptive AI interfaces. *Journal of Cognitive Engineering and Decision Making, 17*(1), 23–37.

Liu, B., & Sundar, S. S. (2018). Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychology, Behavior, and Social Networking, 21*(10), 625–636. https://doi.org/10.1089/cyber.2018.0110

Mahajan, S. (2025). The democratization dilemma: When everyone is an expert, who do we trust? *Humanities and Social Sciences Communications, 12*(1). https://doi.org/10.1057/s41599-025-04734-x

McKee, K. R., Bai, X., & Fiske, S. T. (2023). Humans perceive warmth and competence in artificial intelligence. *iScience, 26*(8), 107256. https://doi.org/10.1016/j.isci.2023.107256

Nguyen, D. S. W., & Shaik, M. M. (2024). Impact of artificial intelligence on corporate leadership. *Journal of Computer and Communications, 12*(4), 40. https://doi.org/10.4236/jcc.2024.124004

Norman, D. A. (1986). Cognitive engineering. In D. A. Norman & S. W. Draper (Eds.), *User centered system design* (pp. 31–61). Lawrence Erlbaum Associates.

Peters, T. M., & Visser, R. W. (2023). The importance of distrust in AI. In *Communications in Computer and Information Science* (p. 301). Springer. https://doi.org/10.1007/978-3-031-44070-0_15

Raees, M., Meijerink, I., Lykourentzou, I., Khan, V.-J., & Papagelis, K. (2024). From explainable to interactive AI: A literature review on current trends in human–AI interaction. *arXiv*. https://doi.org/10.48550/arXiv.2405.15051

Ramachandram, D., Joshi, H., Zhu, J., Gandhi, D., Hartman, L., & Raval, A. (2025). Transparent AI: The case for interpretability and explainability. *arXiv*. https://doi.org/10.48550/arXiv.2507.23535

Riefle, E., & Weber, T. (2022). Adapting AI outputs to user cognitive styles. *Journal of Artificial Intelligence Research, 72*, 157–173.

Riva, G. (2025). The architecture of cognitive amplification: Enhanced cognitive scaffolding as a resolution to the comfort–growth paradox in human–AI cognitive integration. *OSF Preprints*. https://doi.org/10.31219/osf.io/4wcdb_v1

Schmager, S., Pappas, I. O., & Vassilakopoulou, P. (2025). Understanding human-centred AI: A review of its defining elements and a research agenda. *Behaviour & Information Technology*. https://doi.org/10.1080/0144929X.2024.2448719

Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). Customer service chatbots: Anthropomorphism and adoption. *Journal of Business Research, 115*, 14–24. https://doi.org/10.1016/j.jbusres.2020.04.030

Singh, A., Taneja, K., Guan, Z., & Ghosh, A. (2025). Protecting human cognition in the age of AI. *arXiv*. https://doi.org/10.48550/arXiv.2502.12447

Söllner, M., Arnold, T., Benlian, A., Bretschneider, U., Knight, C. L., Ohly, S., Rudkowski, L., Schreiber, G., & Wendt, D. H. (2025). ChatGPT and beyond: Exploring the responsible use of generative AI in the workplace. *Business & Information Systems Engineering*. https://doi.org/10.1007/s12599-025-00932-8

Soomro, R. B., Al-Rahmi, W. M., Dahri, N. A., Almuqren, L., Almogren, A. S., & Aldaijy, A. (2025). A SEM–ANN analysis to examine impact of artificial intelligence technologies on sustainable performance of SMEs. *Scientific Reports, 15*(1). https://doi.org/10.1038/s41598-025-86464-3

Spennemann, D. (2025). The origins and veracity of references ‘cited’ by generative artificial intelligence applications: Implications for the quality of responses. *Publications, 13*(1), 12. https://doi.org/10.3390/publications13010012

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language, 25*(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Srinivasan, T., & Thomason, J. (2025). Adjust for trust: Mitigating trust-induced inappropriate reliance on AI assistance. *arXiv*. https://doi.org/10.48550/arXiv.2502.13321

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

Townsend, B., Plant, K. L., Hodge, J., Ashaolu, O., & Călinescu, R. (2023). Medical practitioner perspectives on AI in emergency triage. *Frontiers in Digital Health, 5*. https://doi.org/10.3389/fdgth.2023.1297073

van Merriënboer, J. J. G., & Sweller, J. (2009). Cognitive load theory in health professional education: Design principles and strategies. *Medical Education, 44*(1), 85–93. https://doi.org/10.1111/j.1365-2923.2009.03498.x

Watermeyer, R., Phipps, L., Lanclos, D., & Knight, C. (2023). Generative AI and the automating of academia. *Postdigital Science and Education, 6*(2), 446. https://doi.org/10.1007/s42438-023-00440-6

Xu, Y., Gao, W., Wang, Y., Shan , X., & Lin, Y.-S. (2024). Enhancing user experience and trust in advanced LLM-based conversational agents. *Computing and Artificial Intelligence, 2*(2), 1467. https://doi.org/10.59400/cai.v2i2.1467

Xiao, Z., Xu, Y., Abdi, N., Collenette, J., Abu-Salma, R., & Sarkadi, Ș. (2024). Banal deception in human–AI ecosystems: A study of people’s perceptions of LLM-generated deceptive behaviour. *arXiv*. https://doi.org/10.48550/arXiv.2406.08386