Association for Information Systems

AIS Electronic Library (AISeL)

Proceedings of the 2023 Pre-ICIS SIGDSA Symposium

Special Interest Group on Decision Support and Analytics (SIGDSA)

12-10-2023

Does AI Metacognitive Ability Lead to Higher AI Advice Utilization?

Bita Mehri
McMaster University, mehrib@mcmaster.ca

Maryam Ghasemaghaei
McMaster University, Hamilton, Ontario, Canada, ghasemm@mcmaster.ca

Follow this and additional works at: https://aisel.aisnet.org/sigdsa2023

Recommended Citation

Mehri, Bita and Ghasemaghaei, Maryam, "Does AI Metacognitive Ability Lead to Higher AI Advice Utilization?" (2023). *Proceedings of the 2023 Pre-ICIS SIGDSA Symposium*. 9.
https://aisel.aisnet.org/sigdsa2023/9

This material is brought to you by the Special Interest Group on Decision Support and Analytics (SIGDSA) at AIS Electronic Library (AISeL). It has been accepted for inclusion in Proceedings of the 2023 Pre-ICIS SIGDSA Symposium by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Does AI Metacognitive Ability Lead to Higher AI Advice Utilization?

Research-in-Progress

Bita Mehri
McMaster University
mehrib@mcmaster.ca

Maryam Ghasemaghaei
McMaster University
ghasemm@mcmaster.ca

Abstract

Metacognition, or "thinking about thinking," is a pivotal ability in human decision-making. As artificial intelligence (AI) evolves, a question emerges: “What occurs when this capability is integrated into AI systems?” This ongoing research investigates the implications of embedding metacognition in AI advisors on user advice utilization. Study 1 assesses users' perceptions of metacognition in AI advisors compared to human experts. Study 2 analyzes how different levels of AI metacognitive abilities (high vs. low) impact AI advice utilization, specifically through perceived AI expertise and system causability. Study 3 gauges how decision type—ill-structured vs. well-structured— and perceived user’s expertise affect interactions with AI metacognition and advice incorporation. Through online experiments, participants interact with advisors, offer initial judgments, receive advice, and make final decisions. Grounded in metacognition theory and the Judge Advisor System (JAS) framework, this research-in-progress aims to elucidate the intricate dynamics between AI metacognition and users' advice utilization.

Keywords

Metacognition, advice utilization, AI advisors, judge advisor system, AI-assisted decision-making.

Introduction

AI, defined as the "science and engineering of making intelligent machines, especially intelligent computer programs (Mesbah et al., 2021), leverages technologies such as machine learning and natural language processing to replicate human cognitive processes. This combination of data and learning models yields practical, data-backed insights that revolutionize decision-making domains such as business forecasting, recruitment strategies, and investment directions (Brynjolfsson & McAfee, 2014). Despite artificial intelligence's remarkable capabilities, recent research consistently underscores users' reluctance to rely on algorithms for tasks traditionally handled by humans, even when these algorithms demonstrate superior performance or promise heightened accuracy and efficiency (Jussupow et al., 2020; Kaufmann et al., 2023). This hesitancy becomes even more pronounced for tasks perceived as subjective due to the common perception that algorithms, while precise, might lack the capacity to comprehend and appreciate nuanced, human-centered decisions (Castelo et al., 2019). A key factor contributing to this aversion is the belief that algorithms, unlike humans, cannot offer the reasoning behind their decisions or exhibit the affective human-like qualities that build trust (Castelo et al., 2019). The ability to fully realize the revolutionary promise of AI technology depends on understanding and overcoming this resistance as businesses around the world attempt to incorporate AI more deeply into their operations.

Considering the challenges associated with the utilization of AI's outputs in decision-making, our research turns to a concept from the realm of human cognitive processes as a potential solution: *metacognition*. Metacognition is simply defined as 'thinking about thinking' (Flavell, 1979). It refers to the awareness and understanding of one's own cognitive processes, which play a pivotal role in our learning and decision-making. More than that, metacognition equips individuals with the ability to plan, monitor, and evaluate their own cognitive strategies, optimizing performance and effectiveness in problem-solving situations (Lai, 2011). Translating the concept of metacognition into the field of artificial intelligence presents a compelling avenue for development. By giving AI systems metacognitive capacities, we could provide them with self-awareness, self-regulation, and self-management capabilities similar to human cognitive processes (Johnson, 2022). As a result, AI systems could be able to provide insights into their internal reasoning processes, evaluate their performance, recognize their limitations, and modify their strategies accordingly (Crowder & Friess, 2011). Such introspective skills might improve the clarity and depth of AI-generated advice, resulting in more robust reasoning that directly addresses one of the previously noted problems with AI-based output utilization. Furthermore, by having AI mimic features of human cognitive behavior, the metacognitive component may reduce the 'otherness' of AI, making its results more familiar and satisfying to human users. In effect, a metacognitive AI could produce advice that is not only more advanced but also more human-like in its presentation and rationale.

This paper seeks to examine the critical role that metacognition may play in bridging the gap between AI systems and human users in dealing with the issues of taking advice from AI. The present research focuses on two proposed mediators: *perceived AI expertise*, which expresses how users evaluate an AI system's competence, and *system causability*, which refers to the extent to which the AI's explanations enable users to understand cause-and-effect relationships in the advice given.

Using the metacognition theory and the judge advisor system, which is a foundational framework that investigates the dynamics of advice-giving and taking (Bonaccio & Dalal, 2006), the primary research questions guiding this study are: *Will users perceive and utilize AI and human advice differently when both exhibit the same level of metacognition (RQ1)? Does enabling metacognition in AI lead to a higher perception of AI expertise and system causability, which consequently may lead to more advice utilization (RQ2)? Moreover, does the decision type or the perceived user's expertise influence the relationship between AI metacognitive ability and advice utilization? (RQ3)* In addressing these research questions, we respond to Kaufmann et al. (2023), who highlighted the need for deeper insights into task-specific algorithm advice acceptance and the understudied role of user domain expertise.

To explore these questions, this research comprises three distinct studies: The initial study seeks to compare advice utilization between human and AI advisors with metacognitive abilities and to ascertain whether metacognition is perceived to be higher in humans or AI. Study 2 aims to examine the effects of AI's metacognition (high vs. low) on AI advice utilization. Study 3 extends the analysis to the interaction between AI's metacognition and various decision types, including well-structured and ill-structured decisions and the impact of the user's self-perceived expertise on the role of metacognition in incorporating AI advice.

The subsequent sections of this research-in-progress paper are organized as follows: In the 'Theoretical Background' section, we draw on both the *metacognition theory* and the *advice utilization literature*, establishing the judge-advisor system as our primary framework. Following this, we present the proposed research model and hypotheses to address the research problems raised. The 'Methodology' section delves into the structure of our three studies, concentrating on the details of the scenario and design of experiments. We conclude the paper by discussing the research's potential implications.

Theoretical Background

Metacognition Theory

Metacognition, as introduced by Flavell in 1979, is essentially "cognition about cognition." However, its scope is significantly broader than this brief description indicates. According to metacognition theory, humans have an in-built ability that grants them the reflective capacity to introspect, understand, and regulate their cognitive processes. Metacognition comprises two major components: *knowledge of cognition* and *regulation of cognition* (Flavell, 1979). Knowledge of cognition refers to a person's awareness of their cognitive ability. This includes an understanding of their own strengths and weaknesses (declarative knowledge), knowledge of problem-solving strategies and processes (procedural knowledge), and an understanding of when and why specific strategies should be applied (conditional knowledge) (Baker, 1989). The second component, regulation of cognition, focuses on managing these cognitive processes. This involves preliminary stages of goal setting and organization (planning), efficient information processing (Information Management), continual self-assessment (monitoring), strategies to identify and correct misunderstandings (debugging), and a post-task review of one's performance (evaluation) (Schraw & Dennison, 1994).

Based on the dual process theory, metacognition is an 'intercession' that can intervene in analytic and experiential processing. Specifically, the dual process theory of cognition proposes two distinct but interwind cognitive systems we use to process information: System 1, characterized by experiential processing, and System 2, associated with analytic processing. System 1 functions automatically, quickly, and largely unconsciously, with little or no effort or thoughtful control. Its efficiency fits familiar situations since it draws extensively on associative memories and experience. Nevertheless, its dependence on past experiences and contextual cues might induce biases, leading to potential errors in reasoning and judgment. Conversely, System 2 operates deliberately and analytically. Its operations are often linked with subjective experiences of agency, choice, and concentration, leading to decisions that arise from a conscious, contemplative, and methodical process, making them less prone to immediate biases (Kahneman, 2011). At its core, metacognitive intercession involves the active and deliberate engagement of awareness and control within the cognitive process. This engagement is particularly heightened during analytic processing, as represented by System 2, due to its inherently conscious nature (Kahneman, 2011). This can lead to the conclusion that relying on metacognitive abilities while decision-making results in less biased and more accurate decisions. Incorporating these theories in the realm of Artificial Intelligence, we posit that an AI system, when instilled with metacognitive capabilities, could potentially harness the essence of System 2 thinking, mirroring human-like, expert cognition. This not only paves the way for more meticulous decisions but, owing to metacognition, AI can also elucidate its cognitive pathways, offering a rationale for its decisions (Flavell, 1979).

Expertise in human cognition is intimately tied to metacognition. Literature in the realm of human cognition reveals that metacognition is more developed in human experts rather than novices. They contend that experts demonstrate stronger self-monitoring skills than their novice counterparts (Chi et al., 2014). In comparative analyses, experts distinctly showcase heightened self-awareness, an attribute rarely echoed by novices (Eteläpelto, 1993). This superior self-knowledge — their acute cognizance of errors, comprehension pitfalls, and task difficulty evaluations — is intimately linked with their domain-specific knowledge, concluding that experts inherently possess amplified metacognitive abilities (Chi et al., 2014). Thus, we state that experts are better at decision-making skills due to their high levels of metacognitive abilities, so enabling AI systems with metacognition might increase users' perceptions of the AI systems' expertise, resulting in more advice utilization.

The literature has argued about the possibilities of embedding metacognition in AI, which traces its roots to the seminal ideas proposed by Marvin Minsky and John McCarthy in the mid-20th century, who promoted the need for machines to represent their knowledge of intelligent behavior declaratively (Cox, 2005). For AI systems, metacognition is intricately defined as the set of processes and mechanisms that enable a computational system to both monitor and oversee its cognitive activities, processes, and structures. This form of control aims to enhance the quality of the system's decisions (Ganapini et al., 2022). Expanding upon this, researchers have proposed a notable SOFAI architecture model, which manifests the principles of Kahneman's "Thinking Fast and Slow" theory. Within this framework, tasks are initially approached by System 1 (S1) solvers, which act swiftly based on previous experiences, akin to human intuitive processes. A subsequent layer, the meta-cognitive (MC) module, assesses the S1 solution, deliberating on the necessity of invoking a deeper, analytical System 2 (S2) solver. This structure, starting with the immediacy of S1 but under the vigilant oversight of the MC agent, strives to harmoniously integrate rapid responses with meticulous analysis (Ganapini et al., 2022). The current body of research suggests that incorporating metacognitive abilities can significantly boost artificial intelligence systems' efficiency, adaptability, safety, autonomy, and awareness of their reasoning (Jackson, 2020; Crowder et al., 2014; Caro et al., 2014). While the literature has extensively explored metacognition's theoretical and conceptual aspects in AI, the emphasis remains predominantly technical. The studies on metacognition in AI primarily offer insights into the mechanisms and architectures, like the SOFAI model or the metacognitive loop (MCL) (Schmill et al., 2008), without delving into empirical evidence that reveals the real-world implications, especially in terms of user perceptions and outcomes in business settings. Our study endeavors to fill this gap, focusing on the impacts of metacognitive AI on user engagement and decision-making.

Advice Taking and Judge-Advisor Systems

The Judge-Advisor System (JAS) is a foundational framework within cognitive sciences that investigates the dynamics of advice-giving and taking (Bonaccio & Dalal, 2006). JAS describes a structured scenario wherein an individual, known as the judge, seeks advice from one or more advisors but retains the ultimate decision-making power (Van Swol, 2011). In the context of AI-assisted decision-making, the user acts as the judge, while the AI system serves as the advisor. Central to understanding advice-taking behavior within this paradigm is the concept of "Weight of Advice" (WOA). WOA is used to measure the level of advice utilization, defined as the extent to which the judge's decisions shift toward the advice given by an advisor (Harvey & Fischer, 1997). A regularly observed trend within the advice-taking literature is the "egocentric discounting" phenomenon. Despite the potential merits of insights from an advisor, judges often shift their initial viewpoints by a minimal 20% to 30% toward the advisor's recommendations (Yaniv & Kleinberger, 2000). This trend can be ascribed to reasons such as the judge's ability to recall their reasoning mechanisms and the lack of understanding into the reasoning of the advisor (Yaniv, 2004b).

Advice utilization is influenced by a myriad of factors, including trust, the advisor's expertise, differences in advice, the judge's proficiency, and the task's nature. These elements can be categorized into four distinct traits: advisor traits, judge traits, task traits, and advice traits (Mesbah et al., 2021). Within the Judge-Advisor System (JAS) literature, the perceived expertise of the advisor is underscored as a pivotal factor in shaping decisions (Mayer et al., 1995). Studies indicate that when an advisor is perceived to possess high competence, judges tend to align their decisions more closely with the advice provided (Schultze et al., 2015). Thus, we posit that enhanced metacognition in AI systems might lead users to perceive them as more expert-like, subsequently amplifying advice utilization.

Moreover, regarding judge characteristics, the judge's perceived expertise—related to one's knowledge, skill, education, and experience—is pivotal in influencing advice acceptance (Kaufmann et al., 2023). Experienced judges or experts in a domain tend to display skepticism towards algorithmic advice, possibly viewing it as a challenge to their professional authority (Logg, 2017). In contrast, individuals with limited expertise or experience, like non-professionals, tend to trust algorithmic suggestions more and are more receptive to incorporating them (Sniezek & Van Swol, 2001). Therefore, we also consider the user's perceived domain expertise here referred to as *self-perceived user expertise*, as a potential factor influencing AI advice utilization.

The Judge-Advisor System (JAS) also acknowledges the impact of task nature in advice taking (Bonaccio & Dalal, 2006). Decision theorists distinguish tasks based on their structure (Simon, 1977). Well-structured decisions, typified by clear criteria, are linear and prioritize computational accuracy. Conversely, ill-structured decisions embrace a cyclical and iterative nature, demanding intricate evaluation (Jonassen, 2010). Simultaneously, advice-taking literature distinguishes between objective, quantifiable tasks, and subjective, intuition-based ones (Castelo et al., 2019; Logg, 2017). Both categorizations highlight that a task's structuredness or objectivity influences advice receptivity. Well-structured or objective decisions tend to increase acceptance of AI recommendations, while ill-structured or subjective decisions prompt users to be more cautious, aligning with findings by Castelo et al. (2019). Hence, we posit that metacognitive AI could differentially impact advice utilization across decision types.

Many studies indicate that rationalizing a recommendation can positively influence users' attitudes about following advice (Ye & Johnson, 1995). However, a prevalent challenge with AI systems is their inherent opacity stemming from the absence of a distinct declarative knowledge representation, making it difficult for these models to provide explanations (Holzinger et al., 2020). This requirement has spurred the development of Explainable AI (xAI). A pivotal notion in the xAI discussion is "*causability*." Holzinger and colleagues proposed the "System Causability Scale" (SCS) aimed at evaluating AI systems' explanation quality (Holzinger et al., 2020). The term "causability" delves into how much an explanation provides a specific level of causal comprehension (Holzinger et al., 2020, p. 196). We, therefore, posit that metacognition, which allows AI systems to supervise their decision-making procedure, will be capable of offering more logical explanations, measurable through system causability.

Research Model and Hypothesis

Based on the JAS framework and drawing from the metacognition literature, we present the research model in Figure 1.

According to prior research, experts were superior to novices in metacognitive knowledge. They consistently demonstrate superior awareness of their cognitive processes, allowing them to effectively monitor their strategies and identify errors during task performance (Eteläpelto, 1993). If AI is enabled with metacognitive ability, it will likely emulate patterns akin to human 'experts.' AI with high metacognition can demonstrate its self-awareness, provide insight into its strategies, and indicate any uncertainties or errors in its processes; users would likely perceive it as possessing a high level of expertise. Moreover, advice utilization literature suggests that one of the critical factors determining whether advice is taken is the perceived competence or expertise of the advice-giver (Bonaccio & Dalal, 2006).

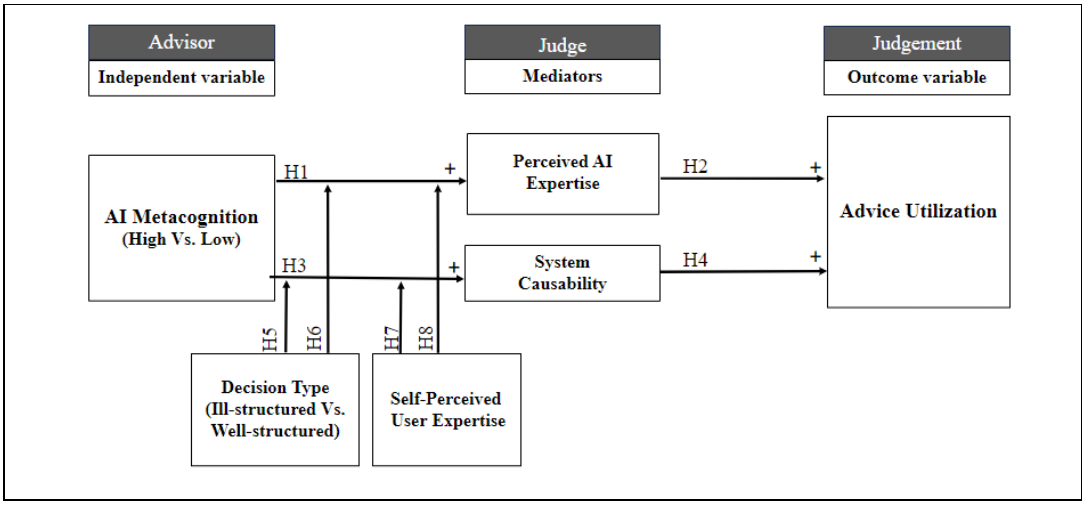

Figure 1. Research Model

Users would be more inclined to trust and utilize its advice or recommendations if an AI system is perceived as more expert due to its metacognitive abilities. The higher the perceived competency of an advisor, the more inclined a judge is to modify their estimate (Schultze et al., 2015). Thus, we hypothesize that:

H1*: AI metacognitive ability will positively influence perceived AI expertise. (Study 2)*

H2*: Higher perceived AI expertise will increase advice utilization. (Study 2)*

The notion of causability revolves around the ability of an entity to elucidate the causal factors and rationale behind its decisions or suggestions. In essence, causability pertains to the explainability and transparency of decisions, enabling a clearer understanding of the "why" behind a decision (Miller, 2019). Based on metacognition literature, metacognitive capabilities will enhance one's ability to articulate the reasoning behind their decisions, as they are acutely aware of their thought processes and can relay this information effectively to others (Koriat, 2012). Additionally, the literature suggests that decision-makers often discount the advisor's opinion relative to their own due to their differential information about the respective underlying justifications. Individuals usually have privileged access to the reasons that lead them to hold their own opinions but only limited access (if any) to the reasons that lead others to hold their opinions. Therefore, the more users understand the cause and effect of the advice suggested to them, the less the disparity between the access to the support of the advisor's opinion and themselves, and the more advice utilization. So, we posit the following:

H3*: AI metacognitive ability will positively impact system causability. (Study 2)*

H4*: Higher System causability will increase advice utilization. (Study 2)*

Decision theorists often categorize decisions into two types based on structuredness: *well-structured* and *ill-structured* (Simon, 1977). Well-structured decisions (e.g., investment decisions), characterized by clear criteria and outcomes, typically prioritize AI's computational power and precision. In such scenarios, while AI's metacognitive abilities might be beneficial, their impact on perceived expertise may not be as pronounced since the decision-making process is straightforward and well-defined (Simon, 1977). On the other hand, ill-structured decision problem-solving (e.g., hiring decision) is not a straightforward process; it is cyclical and iterative and involves evaluation and continuous monitoring (Jonassen, 2010). Ill-structured decisions demand heightened metacognitive and self-regulation skills (Jonassen, 2010). Interacting with AI systems, users might interpret AI's capacity to reflect, adjust, and convey its decision-making processes in uncertain scenarios as a sign of expertise, akin to human experts who navigate complex decisions using similar reflective strategies (Vessey, 1991; Koriat, 2012). Therefore, we hypothesize that:

H5: *Decision type moderates the impact of AI metacognition on perceived AI expertise, with a heightened effect on ill-structured compared to well-structured decisions. (Study 3)*

Moreover, for well-structured decisions marked by clear solution pathways (Simon, 1977), users might lean more toward the algorithmic outputs of AI without profoundly questioning the underlying processes. However, in ill-structured decisions characterized by ambiguity (Vessey, 1991), users will likely emphasize understanding the AI's thought processes. As AI showcases its metacognitive skills in these complex scenarios, it prompts users to delve deeper into the causal reasoning behind AI decisions (Koriat, 2012). Thus, we hypothesize that:

H6: *Decision type moderates the relationship between AI metacognition and system causability, being more pronounced in ill-structured decisions. (Study 3)*

An individual's perceived level of expertise can substantially influence their interaction with and reliance on AI systems. Historically, it is observed that those with more domain-specific expertise, having a nuanced grasp of its intricacies, often evaluate and process AI feedback distinctively than novices (Sniezek & Van Swol, 2001; Harvey & Fischer, 1997). Expert users, equipped with in-depth domain knowledge, evaluate AI systems critically, comparing AI reflections and adjustments against their extensive expertise (Ehrlinger et al., 2008). For them, AI's metacognitive abilities are but one factor among many in gauging its proficiency. Conversely, non-experts, lacking this extensive knowledge, view AI's metacognitive processes as solid indicators of its expertise (Eteläpelto, 1993). Thus, while metacognition may enhance perceived AI expertise for novices, its impact is potentially more muted for self-perceived experts. Furthermore, system causability, the AI's capacity to elucidate its decisions, might also be appraised differently based on self-perceived user expertise. An adept user may dissect the AI's causal explanations more rigorously, delving into its rationale, whereas a novice might value its clarity without probing deeply. Drawing from these insights, we posit that:

H7*: The impact of AI metacognition on perceived AI expertise is weaker for users with higher self-perceived expertise. (Study 3)*

H8*: The influence of AI metacognition on system causability is less pronounced for users with higher self-perceived expertise. (Study 3)*

Methodology

We will use an experimental design to test the research model, manipulating advisor metacognitive ability (Human vs. AI in Study 1; Low vs. High AI metacognition in Studies 2 & 3). While numerous studies have drawn comparisons between AI and human advisors in various dimensions, there remains a significant gap in juxtaposing them based on metacognitive abilities (Castelo et al., 2019). There remains a pressing need to understand how AI's metacognitive abilities compare with those of humans, especially in their influence on advice utilization. Metacognition in humans relates to self-awareness and understanding of one's thought processes. When AI exhibits such abilities, it might be challenging for participants to reconcile this typically human trait with a machine, leading to varying perceptions. To establish a foundational understanding for our subsequent studies, our preliminary study compares the metacognitive ability of humans vs. AI.

The participants will be employees with a basic understanding of technology from different industries in North America who will be recruited with the help of a market research firm. To test the experimental procedures and survey instruments and ensure the constructs' reliability, we will conduct a pilot study involving 50 employees. For the entire survey, the research aims to have a sample size ranging between 100 and 300 participants for each study to ensure statistical significance. However, a power analysis will be conducted to determine the optimal number. To distinctly manipulate the metacognitive ability of AI, participants are introduced to scenarios where AI advisors either demonstrate high metacognition— strong knowledge and regulation of cognition—or portray low metacognition— weak knowledge and regulation of cognition. Interfaces following these scenarios further emphasize the assigned metacognitive levels, allowing participants to form judgments. A manipulation check for AI metacognition will be conducted using an 11-item scale adapted from Urban and Urban (2023). In Study 2, participants will initially review a candidate's job resume and job requirements and rate the candidate's job suitability for a data analyst role (0-100 scale). In Study 3, they will either perform a similar task for a creative team role as an ill-structured decision or will be provided with detailed information about an investment opportunity and decide an investment amount (0-$10,000) as a well-structured decision. Following initial judgments, advice is offered, allowing participants to reevaluate and finalize their decisions. Therefore, advice utilization can be measured objectively via the Weight of Advice (WOA), which is calculated as follows: WOA= (advice–initial estimate)/ (final estimate–initial estimate) (Harvey & Fischer, 1997). In our analysis, we will control for age, familiarity with AI, and trust in the advisor to account for potential confounds. To ensure content validity, measurement scales for the constructs will be selected from the extant literature. Perceived expertise will be measured using a 5-item scale from Ohanian (1990). System causability will be measured using an 8-item scale derived from Holzinger et al. (2020). The user's perceived expertise will be measured using a 4-item scale from Radel et al. (2011).

Conclusion

This study provides several theoretical and practical contributions. Focusing on the perception and utilization of advice from AI with varied metacognitive abilities, incorporating metacognition and advice-taking literature, we developed a research model. We tried to respond to the call by Kaufmann et al. (2023) to delve deeper into the user utilization patterns of AI-based systems. Specifically, we investigated the ripple effects of AI metacognition through the lenses of perceived expertise and system causability, examining their impacts on advice utilization across diverse decision structures and task types. This study endeavors to extend existing literature by delving into the pivotal role of AI metacognition in influencing human decision-making processes. Our exploration seeks to unearth whether and to what extent AI metacognition might influence advice utilization. More crucially, by contrasting well-structured and ill-structured decision environments and examining the role of the user's perceived domain expertise, our research intends to elucidate in which contexts the metacognitive abilities of AI are most beneficial. Regarding practical implications, our findings can guide the design and implementation of AI systems across various sectors. Suppose metacognition in AI does indeed enhance advice utilization. In that case, organizations and developers benefit from designing AI systems that are not only technically adept but also cognizant of human cognitive patterns. The insights from this study can also serve as a beacon for stakeholders when deciding the domains or sectors where AI interventions might be most fruitful, whether in creative arenas or more analytical fields like finance.

References

Baker, L. 1989. "Metacognition, Comprehension Monitoring, and the Adult Reader," *Educational Psychology Review* (1), Springer, pp. 3–38.

Bonaccio, S., and Dalal, R. S. 2006. "Advice Taking and Decision-Making: An Integrative Literature Review, and Implications for the Organizational Sciences," *Organizational Behavior and Human Decision Processes* (101:2), Elsevier, pp. 127–151.

Brynjolfsson, E., and McAfee, A. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, WW Norton & Company.

Caro, M. F., Josyula, D. P., Cox, M. T., and Jiménez, J. A. 2014. "Design and Validation of a Metamodel for Metacognition Support in Artificial Intelligent Systems," *Biologically Inspired Cognitive Architectures* (9), Elsevier, pp. 82–104.

Castelo, N., Bos, M. W., and Lehmann, D. R. 2019. "Task-Dependent Algorithm Aversion," *Journal of Marketing Research* (56:5), SAGE Publications Sage CA: Los Angeles, CA, pp. 809–825.

Chi, M. T., Glaser, R., and Farr, M. J. 2014. *The Nature of Expertise*, Psychology Press.

Cox, M. T. 2005. "Metacognition in Computation: A Selected Research Review," *Artificial Intelligence* (169:2), Elsevier, pp. 104–141.

Crowder, J. A., and Shelli Friess MA, N. C. C. 2011. "Metacognition and Metamemory Concepts for AI Systems," in *Proceedings of the International Conference on Artificial Intelligence (ICAI)*, The Steering Committee of The World Congress in Computer Science, Computer ..., p. 1.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., and Kruger, J. 2008. "Why the Unskilled Are Unaware: Further Explorations of (Absent) Self-Insight among the Incompetent," *Organizational Behavior and Human Decision Processes* (105:1), pp. 98–121.

Eteläpelto, A. 1993. "Metacognition and the Expertise of Computer Program Comprehension," *Scandinavian Journal of Educational Research* (37:3), Taylor & Francis, pp. 243–254.

Flavell, J. H. 1979. "Metacognition and Cognitive Monitoring: A New Area of Cognitive–Developmental Inquiry.," *American Psychologist* (34:10), American Psychological Association, p. 906.

Ganapini, M. B., Campbell, M., Fabiano, F., Horesh, L., Lenchner, J., Loreggia, A., Mattei, N., Rossi, F., Srivastava, B., and Venable, K. B. 2022. "Thinking Fast and Slow in AI: The Role of Metacognition," in *International Conference on Machine Learning, Optimization, and Data Science*, Springer, pp. 502–509.

Harvey, N., and Fischer, I. 1997. "Taking Advice: Accepting Help, Improving Judgment, and Sharing Responsibility," *Organizational Behavior and Human Decision Processes* (70:2), Elsevier, pp. 117–133.

Holzinger, A., Carrington, A., and Müller, H. 2020. "Measuring the Quality of Explanations: The System Causability Scale (SCS) Comparing Human and Machine Explanations," *KI-Künstliche Intelligenz* (34:2), Springer, pp. 193–198.

Jackson, P. 2020. "Toward Metascience via Human-Level AI with Metacognition," *Procedia Computer Science* (169), Elsevier, pp. 527–534.

Johnson, B. 2022. "Metacognition for Artificial Intelligence System Safety–An Approach to Safe and Desired Behavior," *Safety Science* (151), Elsevier, p. 105743.

Jonassen, D. H. 2010. *Learning to Solve Problems: A Handbook for Designing Problem-Solving Learning Environments*, Routledge.

Jussupow, E., Benbasat, I., and Heinzl, A. 2020. *Why Are We Averse towards Algorithms? A Comprehensive Literature Review on Algorithm Aversion*.

Kahneman, D. 2011. *Thinking, Fast and Slow*, macmillan.

Kaufmann, E., Chacon, A., Kausel, E. E., Herrera, N., and Reyes, T. 2023. "Task-Specific Algorithm Advice Acceptance: A Review and Directions for Future Research," *Data and Information Management*, Elsevier, p. 100040.

Koriat, A. 2012. "The Self-Consistency Model of Subjective Confidence.," *Psychological Review* (119:1), American Psychological Association, p. 80.

Lai, E. R. 2011. "Metacognition: A Literature Review," *Always Learning: Pearson Research Report* (24), pp. 1–40.

Logg, J. M. 2017. "Theory of Machine: When Do People Rely on Algorithms?," *Harvard Business School Working Paper Series# 17-086*.

Mayer, R. C., Davis, J. H., and Schoorman, F. D. 1995. "An Integrative Model of Organizational Trust," *Academy of Management Review* (20:3), Academy of Management Briarcliff Manor, NY 10510, pp. 709–734.

Mesbah, N., Tauchert, C., and Buxmann, P. 2021. *Whose Advice Counts More–Man or Machine? An Experimental Investigation of Ai-Based Advice Utilization*.

Ohanian, R. 1990. "Construction and Validation of a Scale to Measure Celebrity Endorsers' Perceived Expertise, Trustworthiness, and Attractiveness," *Journal of Advertising* (19:3), Taylor & Francis, pp. 39–52.

Schmill, M., Oates, T., Anderson, M. L., Josyula, D., Perlis, D., Wilson, S., and Fults, S. 2008. "The Role of Metacognition in Robust AI Systems," in *Workshop on Metareasoning at the Twenty-Third AAAI Conference on Artificial Intelligence*.

Schraw, G., and Dennison, R. S. 1994. "Assessing Metacognitive Awareness," *Contemporary Educational Psychology* (19:4), Elsevier, pp. 460–475.

Schultze, T., Rakotoarisoa, A.-F., and Stefan, S.-H. 2015. "Effects of Distance between Initial Estimates and Advice on Advice Utilization," *Judgment and Decision Making* (10:2), Cambridge University Press, pp. 144–171.

Simon, H. A. 1977. "Scientific Discovery and the Psychology of Problem Solving," in *Models of Discovery: And Other Topics in the Methods of Science*, Springer, pp. 286–303.

Sniezek, J. A., and Van Swol, L. M. 2001. "Trust, Confidence, and Expertise in a Judge-Advisor System," *Organizational Behavior and Human Decision Processes* (84:2), Academic Press, pp. 288–307.

Urban, K., and Urban, M. 2023. "How Can We Measure Metacognition in Creative Problem-Solving? Standardization of the MCPS Scale," *Thinking Skills and Creativity*, Elsevier, p. 101345.

Van Swol, L. M. 2011. "Forecasting Another's Enjoyment versus Giving the Right Answer: Trust, Shared Values, Task Effects, and Confidence in Improving the Acceptance of Advice," *International Journal of Forecasting* (27:1), Elsevier, pp. 103–120.

Vessey, I. 1991. "Cognitive Fit: A Theory-Based Analysis of the Graphs versus Tables Literature," *Decision Sciences* (22:2), Wiley Online Library, pp. 219–240.

Yaniv, I. 2004. "Receiving Other People's Advice: Influence and Benefit," *Organizational Behavior and Human Decision Processes* (93:1), Elsevier, pp. 1–13.

Yaniv, I., and Kleinberger, E. 2000. "Advice Taking in Decision Making: Egocentric Discounting and Reputation Formation," *Organizational Behavior and Human Decision Processes* (83:2), Elsevier, pp. 260–281.

Ye, L. R., and Johnson, P. E. 1995. "The Impact of Explanation Facilities on User Acceptance of Expert Systems Advice," *Mis Quarterly*, JSTOR, pp. 157–172.