Unpacking Human and AI Complementarity: Insights from Recent Works

Yuqing Ren, Xuefei (Nancy) Deng, and K.D. Joshi

Abstract:

In this editorial, we draw insights from recent empirical studies to answer some key questions related to human and AI complementarity. There is consensus regarding the strengths of machine intelligence in performing structured and codifiable tasks and complementing two human limitations: lack of consistency and inability to unlearn conventional wisdom. To work effectively with AI, humans need to possess not only AI skills but also domain expertise, job skills, and metaknowledge to accurately assess human capabilities and AI capabilities. We identify several future directions in understanding the effects of human expertise and experiences on algorithmic appreciation, the mutual learning and adaptions between humans and AI, and the boundary conditions of effective human and AI complementarity.

Keywords: Artificial Intelligence, Human-AI augmentation, Machine Learning, Human Skills, AI-Human Complementarity

Artificial intelligence (AI) is either hailed as a panacea to human-bounded rationality or feared as a harbinger of the demise of humanity. As the steady drumbeat of AI slowly fades away and as we move past the discourse of whether AI is a friend or a foe, it is time for us to embrace the new world where AI and humans coexist. As the dynamic frontier of computational advancements, AI now occupies a prominent role in our society (Berente et al., 2021) and has great potential to augment human capabilities. As Information Systems scholars, we must take a leadership role in understanding the opportunities and challenges of using AI to augment and complement human capabilities.

In this editorial, we borrow insights from recent works to address the following questions around human and AI complementarity: (1) What are the strengths and limitations of human intelligence and machine intelligence? (2) How does the nature of tasks affect human and AI complementarity? (3) What are the optimal ways to structure and coordinate human-AI collaboration? (4) What types of human skills are important for humans to collaborate with AI? (5) How do humans respond to AI, with aversion or appreciation? (6) What affects human willingness to work with AI?

The idea for this editorial originated in a Ph.D. seminar on emerging technologies including AI that the first author taught in the spring of 2023. We started with four articles selected for the seminar, which cover multiple disciplines and domains. We used snowballing to identify a few additional articles. This is an editorial, not a review. Our goal is not to be exhaustive but rather to gather sufficient insights on the aforementioned questions to inform our thinking and serve as inspiration for future research.

Domains, Tasks, and AI Tools

The seven chosen articles cover a broad range of domains including drug discovery (Lou and Wu, 2021), image classification (Fügener et al., 2022), legal predictions (Kleinberg et al., 2018), patent searches (Choudhury et al., 2020), sales calls (Jia et al., 2023), and loan risk prediction (Fu et al., 2021; Ge et al., 2021). Some of the involved tasks are structured, routine, and codifiable such as risk predictions, image classification, conversing with a customer using a script, and searching in a large space, whether for new chemical compounds or relevant prior patents. Other tasks are less codifiable and require the use of common sense or tacit knowledge, such as finding patents that are truly relevant not just textually similar, differentiating between likely drug candidates from spurious ones, or persuading customers to get a credit card.

A wide variety of AI tools were studied in these articles, ranging from automated search tools (Choudhury et al., 2020), to sophisticated machine learning (ML), tree ensemble models such as XGBoost (Fu et al., 2021), gradient-boosted decision trees (Kleinberg et al., 2018), GoogLeNet Inception v3 for image classification (Fügener et al., 2022), and a conversational bot powered by deep learning and natural language processing to inform loan investments (Ge et al., 2021).

Strengths of Machine Intelligence

The consensus from the studies is that AI is generally good at structured tasks, i.e., tasks that are repetitive, codified, and require following pre-specified protocols, but AI is limited in handling unscripted and unstructured tasks (Jia et al., 2023). The former is well suited for AI assistance and the latter makes human involvement not only beneficial but often indispensable. Compared to human endeavors, the use of AI improves efficiency, speed, scalability, consistency, and reliability (Jia et al., 2023). For example, in new drug discovery, AI can screen for new chemical compounds 100 times faster than humans (Lou and Wu, 2021). A reasonably sophisticated ML algorithm has been shown to predict loan listing default probability more accurately than crowd investors, improving the welfare of both investors and borrowers simultaneously (Fu et al., 2021). Rather than using human judges, using AI algorithms to make the bail decision could reduce crime by up to 24.8% with no change in jailing rates or reduce the jail population by 43% with no change in crime rates (Kleinberg et al., 2018).

Why Machines Do Better in Structured Tasks?

Based on the studies, there are at least two reasons why machines outperform humans: accuracy and consistency. This sheds light on two human limitations. The first is humans' overconfidence or tendency to overestimate our abilities (Fügener et al., 2022; Logg et al., 2019). For instance, analyses of bail decisions by human judges suggest that instead of being systemically strict or lenient, human judges make random errors by reacting to "noises" in the data (e.g., defendants' appearances and mood) and mistaking them as "signals" (Kleinberg et al., 2018). In other words, rather than adhering to and applying the rules consistently, human judges try to discern idiosyncratic factors and make corresponding adjustments, believing it would lead to better decisions. In contrast, machines consistently adhere to the rules in the algorithms, resulting in better performance.

The inconsistency of human behaviors also occurs with their decisions to delegate to AI. Due to frequent overestimation of their abilities to perform a task (such as image classification) or underestimation of task difficulty, humans often fail to delegate the task to AI when AI is the better option (Fügener et al., 2022). Even when humans were provided with delegation strategies, which either explained why they should delegate or enforced their delegation, human decisions to delegate difficult images to AI were still quite arbitrary. As a result, although the delegation strategy led to relatively more delegation, there was no significant improvement in decision accuracy (Fügener et al., 2022).

The second human limitation is humans' inability to easily unlearn. As humans gain experience, we acquire knowledge and conventional wisdom to perform various tasks such as investment decisions. While conventional wisdom may apply in traditional financial settings, it may not apply to relatively new settings such as online peer-to-peer (P2P) lending platforms. ML algorithms outperformed crowd investors in predicting the loan default risks on a P2P lending platform because the human investors overweighed the importance of traditional criteria such as credit score, homeownership, and length of credit history (Fu et al., 2021). However, these traditional factors turned out to be poor indicators of borrowers' default likelihood in the P2P lending context because borrowers with higher credit scores, home ownerships, and long credit histories still chose P2P lending (over other funding opportunities) and they tended to be risky and even riskier than those without the good credit scores or history. As humans, we rely on our prior beliefs and struggle to unlearn conventional wisdom. However, machines have no such limitations and can quickly discover or learn new rules in a new setting. Future research needs to examine whether this gives AI a competitive edge over humans in correcting biased behaviors, as machines are unburdened by prior beliefs and hence able to swiftly adapt to new rules, leading to more equitable outcomes.

The Strengths of Human Intelligence

Humans often outperform AI due to domain expertise and tacit knowledge, especially in contexts with incomplete input information. For example, in new patent applications, applicants often use new terms to signal the novelty of their patents and increase their chance of the applications being granted. Hence, AI tools based on textual similarity are not effective at finding truly relevant or "silver bullet" patents. It takes domain knowledge or years of experience for human users to identify additional keywords (e.g., words in the application patent) to increase the chance of finding the silver bullet patents (Choudhury et al., 2020).

There is a caveat in the comparison of human intelligence with machine intelligence though. It is known as the selective labels problem (Kleinberg et al., 2015). Most AI tools today learn from humans. AI algorithms need to be trained on human labeled data after human decisions have been made (e.g., funded loan listings or bail decisions), for which we know the outcome (i.e., default or no default, re-offense or no re-offense). It is called the selective labels problem because only part of all instances are labeled, and human decisions determine which instances have labels. In other words, for the instances that were not chosen by the human judges (e.g., loans unfunded or defendants jailed), we would not know the outcome if the loans had been funded or the defendants had been released. As a result, by default, the machine should achieve higher accuracy than the human (Fu et al., 2021). In other words, machine intelligence builds off human intelligence. Hence it may not be an inherently fair comparison. We should take caution when comparing human and AI performance and advocating for the use of machine intelligence to replace human intelligence.

Labor Division between Humans and AI

We can achieve better performance by combining human intelligence and machine intelligence and leveraging the strengths of both while mitigating their respective weaknesses. The question then becomes whether there is an optimal way to structure and coordinate the work of humans and AI. Some of the studies suggested a sequential division of labor with AI starting the structured work (e.g., telling customers about a credit card product by following a script) and delegating less structured tasks (e.g., answering customers' out-of-knowledge-bank questions about the product) to humans (Jia et al., 2023). In the context of image classification, such an arrangement (AI delegation to human) has been shown to outperform human delegation to AI, AI alone, or humans alone (Fügener et al., 2022). A key reason why AI delegation (also known as AI inversion) is more effective than human delegation is the weak metaknowledge of humans. Humans' inability to accurately assess their own capabilities or gauge the difficulty of a task results in their failure to delegate to AI when the delegation would have been appropriate.

Human Skills to Benefit from AI Assistance

Not all humans benefit equally from AI assistance. Studies have identified four human factors – humans' AI and technical skills, domain expertise, job experience, and metaknowledge – as important for human and AI complementarity. First, knowledge about AI systems and AI skills are important for humans to work effectively with AI. It may become an essential skill for future generations of professionals, similar to computer skills in recent decades. Technical skills or computer science knowledge is also important. Having technical skills not only allowed human workers to use AI tools more productively (e.g., manipulating a word cloud by adding relevant keywords) but also enabled them to derive greater benefits from AI tools (Choudhury et al., 2020).

Second, domain expertise is essential for human and AI complementarity. Studies of firms' AI capabilities found that a firm's AI innovation capability comes from employees possessing a combination of AI skills and domain expertise (Lou and Wu, 2021). Domain expertise is as crucial as AI skills because new drug discovery requires an iterative process of searching a large space for new chemical compounds and differentiating the drug candidates from spurious ones.

The third crucial human factor is job skills or job experiences. Higher-skilled workers benefit more from AI complementarity than lower-skilled workers. When a sales firm adopted a conversational bot to assist human representatives with sales calls, AI assistance helped human workers by screening out uninterested customers so that human workers can focus on conversing and persuading more serious customers (Jia et al., 2023). Interestingly, AI assistance had a much greater positive impact on higher-skilled employees than lower-skilled employees, by enabling higher-skilled workers to answer more customer questions that are outside of the knowledge bank, make more sales, and experience better mood, higher morale, and greater freedom (Jia et al., 2023). In contrast, lower-skilled workers struggled to solve untrained questions and reported greater stress, a stronger sense of defeat, and lower morale (Jia et al., 2023).

Finally, several studies have shown the importance of humans' metaknowledge or metacognition when working with AI. Like humans, AI makes mistakes, and humans can catch and compensate for AI errors. When physicians used AI tools to assist with medical diagnosis, correct AI advice improved diagnosis accuracy and incorrect AI advice reduced accuracy. AI augmentation is most successful when human decision makers have accurate assessments of both their own capabilities and AI capabilities, i.e., metacognitions of self-monitoring and system monitoring (Jussupow et al., 2021). In another study using AI tools for radiology diagnosis (Lebovitz et al., 2022), medical professionals faced uncertainty when AI results differed from their judgements. Under uncertain circumstances, professionals who engaged in AI interrogation practices by relating their own knowledge to AI knowledge benefited more from AI advice than those who failed to engage in such practices (Lebovitz et al., 2022).

Algorithmic Aversion or Algorithmic Appreciation

The last set of questions we explore are regarding human willingness to use AI assistance. For optimal performance, it is essential for humans to be willing to delegate to or collaborate with AI, allowing both to complement each other and achieve high performance. The studies provide some high-level insights into the debate between algorithmic aversion and algorithmic appreciation. Despite the general consensus among researchers on algorithmic aversion, Logg et al. (2019) uncovered robust experimental evidence to show that non-experts adhered more to advice from an algorithm than from a person in performing the tasks of guessing human weight from a picture, forecasting song popularity, and predicting romantic attraction. Meanwhile, they also showed that algorithm appreciation weakened when subjects chose between an algorithm's decision and their own. Even in these cases, subjects discounted algorithmic advice less than they discounted advice from other humans (Logg et al., 2019).

Human users vary in their degree of algorithmic appreciation or aversion. Two human factors – expertise and past performance – have been linked to one's willingness to follow AI advice. Experts who have the knowledge or experience with what the algorithms do (e.g., geopolitical forecasting) relied less on algorithmic advice than nonexperts; as a result, the experts' adherence to their prior judgments and their failure to utilize machine intelligence ultimately lowered their forecast accuracy relative to those nonexperts (Logg et al., 2019). Moreover, human's willingness to follow AI advice is likely to be affected by their past performance. Human investors who had encountered more defaults in their manual investing were less likely to adopt a Robo Advisor designed to assist with investment decisions (Ge et al., 2021). These findings reveal an interesting problem, that is, humans who potentially can benefit more from working with AI tend to be less willing to appreciate and follow AI advice. Hence, the exact effects of human expertise and experience on human-AI collaboration performance are contingent upon the domain and labor division between the two (e.g., AI delegating to humans or vice versa).

Advancing Research on Human and AI Complementarity

To summarize, our analysis of recent studies on human and AI complementarity reveals a growing consensus regarding the strengths and limitations of machine intelligence and human intelligence. It also identifies productive ways for dividing labor and fostering collaboration between AI and humans. Furthermore, certain human skills coupled with the appreciation of AI are key to working effectively with AI.

One unexplored and promising future research direction is the mutual learning and adaptation between humans and AI. Most existing studies regarded human intelligence and machine intelligence as fixed, at least during the studied period. In practice, both are constantly evolving and learning from one another. Human labels are used as ground truth to train AI algorithms and ideally, humans can learn from AI advice and explanations. Both the source and the knowledge used to train AI tools can influence their performance. For instance, Lebovitz et al. (2021) found that AI tools trained with know-what knowledge did not meet expectations and the issue can be addressed by training AI with rich know-how knowledge. Hence, future research should continue to explore how best to use human data to train AI tools, how to design AI tools to improve human learning, and the dynamic, iterative interplay between the two.

Due to the limited number of empirical studies that we consulted, future research should continue testing the boundary conditions of our observed patterns and doing so in a broader set of contexts. By diversifying the contextual elements (such as domains, tasks, and labor divisions), researchers can investigate intricate nuances and contingencies that underlie human and AI complementarity. Such a contingent approach is essential for deepening our understanding of human and AI complementarity forward.

References

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, 45(3), 1433-1450.

Choiudhary, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal,* 41(8), 1381-1441.

Fu, R., Huang, Y., & Singh, P. (2021). Crowds, lending, machine, and bias. *Information Systems Research,* 32(1), 72-92.

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2022). Cognitive challenges in human–artificial Intelligence collaboration: Investigating the path toward productive delegation. *Information Systems Research*, 33(2), 678-696.

Ge, R., Zheng, Z., Tian, X., & Liao, L. (2021). Human–robot interaction: When investors adjust the usage of robo-advisors in peer-to-peer lending. *Information Systems Research*, 32(3), 774-785.

Jia, N., Luo, X., Fang, Z., & Liao, C. (2023). When and how artificial intelligence augments employee creativity. *Academy of Management Journal*. Published Online:28 Mar 2023. https://doi.org/10.5465/amj.2022.0426

Jussupow, E., Spohrer, K., Heinzl, A., & Gawlitza, J. (2021). Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. *Information Systems Research*, 32(3), 713-735.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133(1), 237-293.

Lebovitz, S., Levina, N., & Lifshitz-Assaf, H. (2021). Is AI ground truth really 'true'? The dangers of training and evaluating AI tools based on experts' know-what. *MIS Quarterly*, 45(3), 1501-1525.

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organization Science*, 33(1), 126-148.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90-103.

Lou, B., & Wu, L. (2021). AI on drugs: Can artificial intelligence accelerate drug development? Evidence from a large-scale examination of bio-pharma firms. *MIS Quarterly*, 45(3), 1451-1482.