INTERNATIONAL DESIGN CONFERENCE – DESIGN 2024
https://doi.org/10.1017/pds.2024.227

Human-AI collaboration by design

Binyang Song [1,✉], Qihao Zhu [2] and Jianxi Luo [2]

[1] Virginia Tech, United States of America, [2] Singapore University of Technology and Design, Singapore

✉ binyangs@vt.edu

Abstract

Human-AI collaboration (HAIC) is a promising strategy to transform engineering design and innovation, yet how to design artificial intelligence (AI) to boost HAIC remains unclear. Accordingly, this paper provides a new, unified, and comprehensive scheme for classifying AI roles. On this basis, we develop an AI design framework that outlines expected AI capabilities, interactive attributes, and trust enablers across various HAIC scenarios, offering guidance for integrating AI into human teams effectively. We also discuss current advancements, challenges, and prospects for future research.

Keywords: collaborative design, artificial intelligence (AI), innovation

1. Introduction

The concept of 'Artificial Intelligence' (AI) was first conceived in the 1950s with the aspiration of machines exhibiting human-like intelligence (Turing, 1950). Since its inception, AI has progressed from early theoretical concepts to rule-based expert systems, advancing further into machine learning and deep learning. Recently, we have entered the era of artificial general intelligence (AGI), marked by the rapid rise of large language models (LLMs) such as Generative Pre-trained Transformers (GPT). State-of-the-art AI has unlocked various applications, particularly in generative capabilities within design innovation (Luo, 2022). For instance, DALL-E 3 (Ramesh et al., 2021) and Stable Diffusion (Rombach et al., 2022) can generate high-quality images from complex text descriptions, and have been applied in industrial design (Liu and Hu, 2023). The latest AI models, including point-E (Nichol et al., 2022) and Dream Fusion (Lan, 2022), can produce three-dimensional (3D) shapes from text prompts, holding significant potential to revolutionize engineering design and manufacturing. AGI models like ChatGPT (OpenAI, 2023) excel at tasks such as question-answering and information summarization, which have proven effective in enhancing novelty and usefulness in concept generation (Filippi et al., 2023).

As AI integrates into the workforce, a burgeoning debate centres on AI's role at work. Humans and AI possess unique and complementary strengths: humans bring creativity, emotional intelligence, generalization, and ethical decision-making, while AI boasts computational power, resulting in high-speed and scalable data processing, and the ability to perform both repetitive and creative generation tasks. Recent consensus between academia and industry suggest that incorporating AI into human teams is a promising strategy for transformative outcomes (Vorobeva et al., 2023; Luo, 2023). The Deloitte Institute has categorized the evolving relationship between humans and AI into three stages: 'substitution,' where technology automates tasks previously done by workers; 'augmentation,' where technology assists workers, empowering transformation for greater value; and 'collaboration,' where technology and workers jointly innovate, creating meaningful transformation and driving gains in cost, efficiency, and value (Deloitte, 2020). Similarly, other researchers consider AI's societal impact from three perspectives: technology-centric, human-centric, and collective intelligence-centric,

https://doi.org/10.1017/pds.2024.227 Published online by Cambridge University Press

resonating with the three stages (Peeters et al., 2021). Across this spectrum, AI assumes various roles in human-AI collaboration (HAIC), requiring varying AI capabilities and human-AI interaction patterns. While AI continues evolving, it is unclear how to craft AI capabilities and interactive attributes for particular HAIC collaboration scenarios, such as collaboration for design innovation.

To address this gap, we aim to answer: How to classify AI roles and design AI accordingly to foster HAIC? By developing a framework to guide the design of AI, our contributions are twofold: (1) We introduce a new scheme for classifying AI roles in HAIC that is both unified and comprehensive, capable of distinctly classifying any AI use case, reflecting the full scope of an AI agent's abilities. (2) Building on this, we develop an AI design framework that delineates the expected AI capabilities, interactive attributes, and trust enablers across various HAIC contexts from multiple perspectives. This framework offers a guideline for informing AI design, promoting the continued integration of AI into human teams for engineering design, innovation, and wider applications.

The paper is structured as follows: Section 2 reviews the background of HAIC and the roles of AI in human-AI hybrid teams. Section 3 introduces the new scheme for classifying AI roles. Section 4 proposes and discusses the design of AI in various HAIC settings. The paper concludes in Section 5 by underscoring its contributions and limitations.

2. Literature review

In this section, we examine the efficacy and framework design of HAIC, as well as the roles of AI across diverse application scenarios.

2.1. Human-AI collaboration

The nature of problems across many domains is evolving to become highly knowledge-intensive, interdisciplinary, and complex, surpassing the capabilities of individual humans and specialized AI (Memmert and Bittner, 2022). HAIC emerges as a promising paradigm to harness the complementary strengths of humans and AI for problem-solving, insight generation, and value creation. It has been referred to in various contexts as hybrid intelligence (Dellermann et al., 2019), hybrid human-AI teaming (Caldwell et al., 2022), and superteams (integration of AI into teams) (Deloitte, 2020).

Despite its potential, the interdisciplinary socio-technological field of HAIC is rife with unanswered questions, presenting high risks when misapplied. Prior studies in engineering design and innovation have reported mixed successes of AI in improving team performance. On one hand, AI has proven helpful in some instances, expediting designer learning (Viros-i-Martin and Selva, 2022), enhancing design performance at individual and team levels (Song and Zurita et al., 2022; Song and Gyory et al., 2022), boosting analytic and decision-making abilities (Chong et al., 2023), elevating creativity (Song et al., 2021), improving team coordination (Gyory et al., 2021), and strengthening team agility (Song and Gyory et al., 2022). On the other hand, some studies show that AI may not always be beneficial, indicating it is not a universal solution for design problems (Chong et al., 2022), can hinder high-performing teams (Zhang et al., 2021), or negatively affect the learning process of designers (Viros-i-Martin and Selva, 2019).

To guide the design and framing of human-AI hybrid teams, a strand of research focuses on developing comprehensive frameworks and identifying key design areas of HAIC. Dellermann et al. (2019) proposed a framework categorized into four dimensions, including task characteristics, learning paradigms, and human-AI interactions, each with several sub-dimensions. Dubey et al. (2020) presented a similar structure, substituting human-AI interactions with teaming characteristics and trust. Seeber et al. (2020) developed a research agenda to explore the potential risks and benefits of HAIC, proposing three design dimensions - AI artifact, collaboration, and institution—each with accompanying research questions. Figure 1 visualizes a framework synthesized from these prior studies (Dellermann et al., 2019; Dubey et al., 2020; Seeber et al., 2020), depicting the design areas of HAIC. The top part describing the facets of tasks, such as task type, goal, allocation, and role of AI in HAIC. The middle details the learning paradigm between humans and AI, including the mutual learning and augmentation. The bottom illustrates human-AI interactions like AI autonomy, interactive attributes, trust enablers, and information flow.

https://doi.org/10.1017/pds.2024.227 Published online by Cambridge University Press

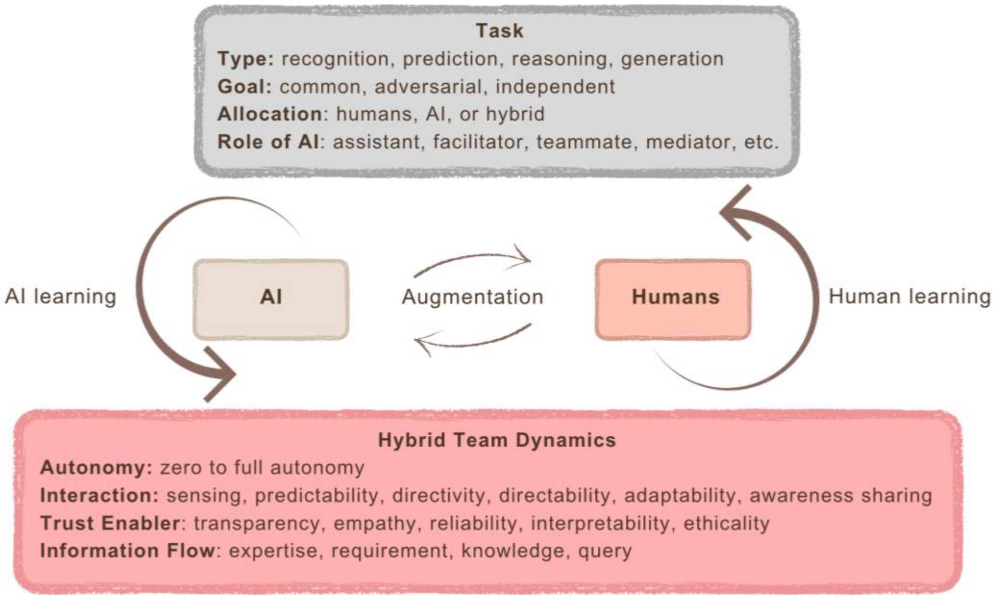

Task
Type: recognition, prediction, reasoning, generation
Goal: common, adversarial, independent
Allocation: humans, AI, or hybrid
Role of AI: assistant, facilitator, teammate, mediator, etc.
AI learning
AI
Augmentation
Humans
Human learning
Hybrid Team Dynamics
Autonomy: zero to full autonomy
Interaction: sensing, predictability, directivity, directability, adaptability, awareness sharing
Trust Enabler: transparency, empathy, reliability, interpretability, ethicality
Information Flow: expertise, requirement, knowledge, query

Figure 1. The human-AI collaboration framework employed in this paper

2.2. The roles of AI

The intended role of AI is a pivotal design choice in HAIC. Various systematic approaches to categorizing AI roles have been proposed in previous research. For instance, Bruemmer et al. (2002) suggested that AI's potential roles range from a tool controlled by humans to a subordinate conducting high-level tasks under minimal supervision, an independent equal, and a leader overseeing others in performing tasks. Dubey et al. (2020) identified four roles: task-oriented personal assistant, coordination-oriented teamwork facilitator, cognitively able human-like associate, and collective moderator. Based on expert interviews, Siemon (2022) defined roles such as the coordinating leader, the creative idea generator, the detail-oriented perfectionist, and the practical doer. Bittner et al. (2019) introduced a taxonomy with roles including the facilitator, peer, and expert. Other researchers have delineated roles with greater granularity, such as the student idea summarizer, evaluator, mediator, arbitrator, and tutor across detailed application scenarios (Tan et al., 2023). Furthermore, Papachristos et al. (2021) found that humans may perceive the same AI differently, as a confirming mirror, a helpful assistant, a guiding advisor, or a trusted oracle. While there is overlap among these AI role schemes, a unified and comprehensive classification system that can distinctly categorize all AI roles is lacking.

As AI assumes different roles in HAIC, users may exhibit varied cognitive, emotional, and behavioral responses, influencing their trust toward and acceptance of AI (Vorobeva et al., 2023). In the current development phase, utilizing AI to assist and augment humans, rather than replacing them, can enhance enjoyment, ease of use, and overall acceptance of AI (Vorobeva et al., 2023). While the intended role of AI determines its expected attributes in HAIC, the boundaries among different AI roles are ambiguity in current AI role defining systems, which is not informative for framing human-AI hybrid teams.

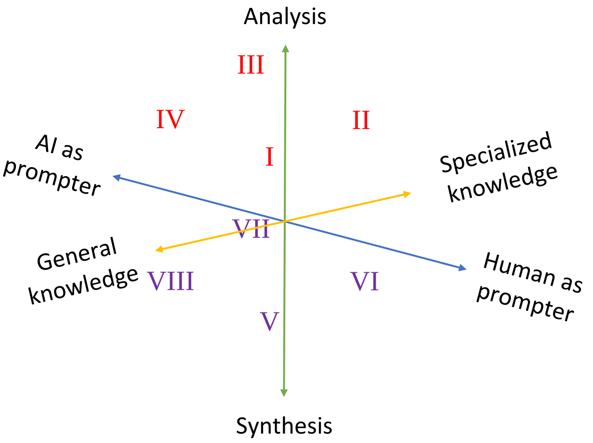
To address the gap, this paper aims to develop an inclusive, informative, and unified scheme for defining AI roles, which is inspired by the reviewed schemes but can effectively inform the design of HAIC. On this basis, we further discuss how the role of AI influences three particular design areas of AI development, namely expected capabilities, interactive attributes, and trust enablers, in the HAIC framework reviewed in this section. In this paper, the classification scheme for AI roles and the design of AI for HAIC draw inspiration from the literature reviewed in this section, as well as from the authors' first-hand experiences with advanced Large Language Models (LLMs) and their extensive background in developing and interacting with various AI agents.

3. A New AI role definition scheme

In this section, we propose a new scheme that offers a unified and comprehensive classification of AI roles. This scheme is delineated by three dimensions: Initiation Spectrum (Human as Prompter vs. AI

https://doi.org/10.1017/pds.2024.227 Published online by Cambridge University Press

as Prompter), Intelligence Scope (Specialized vs. General), and Cognitive Mode (Analysis-Oriented vs. Synthesis-Oriented). These dimensions classify AI roles into eight categories (2×2×2), as illustrated in Figure 2.



I. Analysis + General + Human (p): *search engine, consultant, etc.*
II. Analysis + Domain-specific + Human (p): *project tracker, knowledge manager, etc.*
III. Analysis+ Domain-specific +AI (p): *analyst, recommender, etc.*
IV. Analysis + General + AI (p): *suggestor, etc.*
V. Synthesis + General + Human (p): *ideator, generator, etc.*
VI. Synthesis + Domain-specific + Human (p): *expert system, etc.*
VII. Synthesis + Domain-specific + AI (p): *facilitator, mediator, etc.*
VIII.Synthesis + General + AI (p): *adversary, innovator etc.*

Figure 2. The proposed scheme for classifying AI roles in human-AI collaboration

3.1. Initiation spectrum: Human as Prompter - AI as Prompter

This 'Initiation Spectrum' dimension, 'Human as Prompter - AI as Prompter,' examines the interaction dynamics between humans and AI in collaborative settings. At one end is the 'Human as Prompter' scenario, where humans actively guide and direct the AI by setting parameters, providing datasets, or issuing textual prompts, reflecting a traditional view of AI as a tool or assistant responding to explicit human instructions. Here, the AI's role is primarily reactive, relying on the human partner for direction.

At the other end is 'AI as Prompter,' where the AI takes a proactive role, exemplified by recommender engines on online platforms that analyse user behaviour to provide targeted suggestions autonomously. In this case, the AI independently gathers and processes information, often in real-time, to assist or guide human decisions, marking a shift towards AI systems that can potentially anticipate needs and offer solutions without direct human prompting, thus becoming more integral participants in the collaboration.

In summary, the 'Human as Prompter - AI as Prompter' dimension critically examines the source and nature of inputs in AI-human interactions, highlighting a spectrum from human-dominated to AI-initiated collaboration.

3.2. Intelligence scope: Specialized - General

The “Intelligence Scope” dimension, 'Specialized - General,' assesses the scope of the knowledge base underpinning AI, ranging from narrow, domain-specific to broad, general applications. A specialized knowledge base is typified by an intense focus on particular domains and tasks. This is common in design innovation, where AI systems are engineered for precise tasks within specific product domains, such as predictive modelling in certain engineering fields or market analysis in business contexts. Nevertheless, the advent of Large Language Models (LLMs) is altering this landscape, propelling us toward Artificial General Intelligence (AGI). Models like GPT-4 are emblematic of this evolution, showcasing the ability to operate across a wide array of domains with a broad and general knowledge base, indicating more versatile and adaptable AI capabilities (Bubeck et al., 2023).

This dimension, influenced by the framework of Zhu & Luo (2023a), classifies an AI agent's capabilities in terms of knowledge and reasoning. It recognizes that AI intelligence exists along a continuum within this dimension. For instance, a generative AI designed for topology optimization under specific conditions may be trained on a specialized dataset, limiting its knowledge base and reasoning to a narrow field. Conversely, general generative AI models like DALL-E 3 are trained on diverse datasets, enabling them to obtain a broad knowledge base with cross-domain reasoning abilities.

Overall, this dimension evaluates the spectrum of an AI's knowledge scale, ranging from highly specialized to broad, cross-domain capabilities.

https://doi.org/10.1017/pds.2024.227 Published online by Cambridge University Press

3.3. Cognitive model: Analysis-oriented - Synthesis-oriented

The “Cognition Mode” dimension, “Analysis-oriented – Synthesis-oriented”, indicates the target functions of AI. Analysis in the AI context is the process of breaking down and interpreting complex data. It involves a variety of prediction tasks such as pattern recognition, classification, and regression. In the field of design innovation, AI's role in analysis is critical and multifaceted. It begins with understanding user needs, which includes analysing online reviews to decipher consumer preferences and insights (Siddharth et al., 2022). This extends to retrieving technical knowledge through knowledge-based systems (Siddharth et al., 2022; Luo et al., 2019), aiding informed decision-making in design. Later, AI plays a crucial role in the automatic evaluation of product or prototype functional performance (Song et al., 2023), ensuring that design outputs align with desired specifications. The analytical approach is predominantly reductive, aiming to distil complex data into more manageable and meaningful forms for the innovation process, such as patterns, indicators, or labels, to facilitate understanding and application.

Synthesis, in contrast, focuses on the creation and generation of new information, insights, solutions, or designs. This encompasses tasks like concept generation or shape synthesis, which are essential in innovation and design. AI's role in synthesis within design innovation has traditionally been significant in the later stages. For example, AI is utilized in topological optimization and generative design to create and optimize design solutions for improved functional performance (Regenwetter et al., 2022). However, the scope of AI in synthesis is broadening, with recent advancements applying it to earlier stages of design, including functional concept generation (Zhu & Luo, 2023a; Zhu et al., 2023) and potentially developing empathetic understanding of stakeholders (Zhu & Luo, 2023b), thereby expanding AI's application in design. Synthesis, as opposed to analysis, does not extract or simplify complex data but rather integrates and enriches it, transforming it into new understandings, innovative solutions, and ideas.

This dimension is critical as it delineates AI's orientation in processing data: whether it leans towards the extraction or integration of information.

4. Design of AI for Human-AI collaboration

Following the proposed scheme, we provide a guideline for framing the design of AI for HAIC in terms of tasks, interactive attributes, and trust enablers of AI in this section.

4.1. Capabilities of AI

Specifying AI capabilities is crucial for AI design as it directs focused development, ensures efficient resource allocation, optimizes performance, and aligns the AI agent with ethical standards and user expectations. Dellermann et al. (2019) classified AI tasks into four categories: recognition, prediction, reasoning, and action. In the context of engineering design and innovation, we replace 'action' with 'generation'. Table 1 defines each capability and summarizes the AI capabilities (columns) expected in various HAIC settings (rows). For instance, *recognition* involves AI recognizing patterns, objects, or concepts from data. According to the recognition column, this capability is expected when a human or an AI agent is the prompter for understanding prompts or context; it is anticipated when the AI agent is analysis-oriented; it is optional when the AI agent is synthesis-oriented, only necessary to identify concepts from input guidance for guided synthesis; it is influenced by the knowledge basis underpinning the AI agent, with a broader knowledge basis enabling more complex cross-domain recognition. According to the H row (i.e., the first row under the title row), a human being the prompter requires AI capabilities of recognition, prediction, and reasoning for understanding human prompts, which does not determine if the generation capability is needed. To ascertain which AI capabilities are expected for a given HAIC setting, we need to integrate multiple rows. For example, under the “Initiation Spectrum: human (H) – Intelligence Scope: specialized (Sp) – Cognition Mode: synthesis (S)” setting, by combining the H, Sp, and S rows, we can deduce that all AI capabilities are expected for one or multiple purposes. Given a collaboration setting, Table 1 can serve as a guideline for designing the capabilities of the corresponding AI agent.

https://doi.org/10.1017/pds.2024.227 Published online by Cambridge University Press

4.2. Interactive attributes of AI

AI interactive attributes are important because they enable AI systems to communicate, collaborate, and adapt effectively in dynamic environments, enhancing their utility, efficiency, and user experience. This makes AI more accessible, relevant, and valuable in real-world applications. The interactive attributes considered in our framework are mainly adapted from those developed by Dubey et al. (2020) and Seeber et al. (2020). Table 2 defines each interactive attribute and summarizes the AI interactive attributes (columns) expected in various HAIC settings (rows). The prompter in a hybrid human-AI team actively initiates communication, taking a leading role in guiding the collaborative problem-solving process. In our framework, we focus on different initiation settings to discuss AI's interactive attributes. When a human is the prompter, the AI agent is expected to exhibit directability, allowing it to be guided by human input or predefined algorithms and granting the human prompter greater autonomy to control or adjust the AI agent. Conversely, when the AI agent is the prompter, it should demonstrate sensing ability, predictability, directivity, adaptability, and awareness-sharing ability, to actively detect, predict, and adapt to contexts, request human complementation, and share insights for better collaboration. These interactive attributes endow the AI prompter with more human-like cognitive abilities to guide the collaborative endeavour.

Table 1. Expected capabilities of AI under various human-AI collaboration settings: H - human, Sp - specialized, G - general, A - analysis, S - synthesis

		Recognition: *recognize patterns, objects, or concepts from data*	**Prediction**: *analyse historical data and draw patterns to forecast future*	**Reasoning**: *process information, draw inferences, and make decisions*	**Generation**: *create new content / designs by combining existing elements*
Initiation Spectrum	*H*	Expected for human prompt understanding			-
	AI	Expected for context understanding			
Intelligence Scope	*Sp*	Affected by knowledge basis: a broader knowledge basis enables more AGI, supporting cross-domain recognition, prediction, reasoning, and generation			
	G				
Cognition Mode	*A*	Expected for all kinds of analysis tasks		Expected to understand patterns and contexts, apply learned rules to make inferences (A) / combine elements for generation (S)	Optional
	S	Optional to enable guided synthesis by identifying concepts from input guidance	Optional to enable guided synthesis by evaluating synthesized samples		Expected for all kinds of synthesis tasks

Table 2. Expected interactive attributes of AI under various human-AI collaboration settings: H - human, Sp - specialized, G - general, A -analysis, S - synthesis

		Sensing: *perceive and interpret real-world data to interact with and understand environment*	**Predictability**: *discerning and understanding future trends, intentions, and activities*	**Directivity:** *direct the attention of humans to critical features, suggestions, and warnings*	**Directability:** *be guided or controlled by humans or specific programming*	**Adaptability**: *adjust its behaviour or algorithms in response to changes in its environment or data*	**Awareness sharing***: communicate and share insights or data-driven understanding with humans*
Initiation Spectrum	*H*	Optional	Optional	Optional	Expected to be responsive to external guidance	Optional	Optional
	AI	Expected to detect human statuses and problem conditions	Expected to infer human status and problem condition evolution	Expected to direct human attention to obstacles that AI encounters	Optional	Expected to actively adapt to varying contexts	Expected to facilitate collaborative decision-making and actions
Intelligence Scope	*Sp*	-					
	G						
Cognition Mode	*A*						
	S						

https://doi.org/10.1017/pds.2024.227 Published online by Cambridge University Press

4.3. Trust enablers

In HAIC, humans interact with AI socially, where trust is an attitude held by humans that AI can help solve problems featuring uncertainty and vulnerability. Trust is vital to ensure safety and reliability, particularly in high-stakes environments where AI decisions can have significant consequences. It fosters user adoption and engagement, as people are more likely to use and benefit from AI they trust (Schelble et al., 2022). Additionally, trust is essential for ethical decision-making and effective collaboration, as it builds confidence in AI's ability to handle tasks, learn from interactions, and respect privacy and fairness. In this paper, we refer to *trust enablers* as a set of AI attributes that can foster the attitude of trust toward AI and should be considered for developing collaborative AI. Table 3 lists the definition of each trust enabler and summarizes the AI trust enablers (columns) expected in various HAIC settings (rows). Among all trust enablers, empathy, as an interaction-related enabler, is only considered in terms of the initiation spectrum dimension, whereas all others are considered from all three dimensions, each offering a distinct perspective to design the trust enablers. For example, to design AI's transparency in HAIC, we need to consider three perspectives, which inform humans about the limitations in human prompt or context detection and interpretation in the Initiation Spectrum dimension, the potential biases in data from which the knowledge bases are learned in the Intelligence Scope dimension, and the limitations of the processes followed during the development and deployment of the AI agent in the Cognition Mode dimension. Table 3 can be used as a guideline to design AI trust enablers by following one perspective or integrating multiple perspectives.

Table 3. Expected trust enablers of AI under various human-AI collaboration settings: H - human, Sp - specialized, G - general, A -analysis, S - synthesis

		Transparency: *be open and clear about how AI systems are designed, developed, and deployed*	**Empathy**: *recognize, interpret, and respond to human emotions in a human-like manner*	**Reliability**: *perform consistently and accurately across conditions and over time*	**Interpretability**: *explain or provide the reasoning behind a specific decision or output produced by the AI*	**Ethicality**: clarify *responsibility, privacy, fairness, safety, long-term implications, social rule compliance*
Initiation Spectrum	*H*	Expected to inform about the limitations in the detection and interpretation of human prompts (H) / context (AI)	Expected to capture and respond to human emotions conveyed by prompts (H) / context (AI) in varying conditions	Expected to evaluate how much AI adapts to changing conditions	Expected to explain how much outputs are correlated with human prompts (H) / contexts (AI)	Expected to protect privacy, respect human consent, and determine who is accountable for AI failure
	AI					
Intelligence Scope	*Sp*	Expected to inform about the potential biases in data from which the knowledge bases are learned	-	Expected to evaluate when and how much to trust the corresponding training data and learned knowledge basis	Expected to explain how knowledge elements interact for reasoning	Expected to identify, mitigate, and prevent biases in training data
	G					
Cognition Mode	*A*	Expected to inform about the limitations of the processes followed during development and deployment		Expected to evaluate when and how much to trust the corresponding functionality	Expected to explain how input data is processed and transformed into a prediction	Expected to mitigate issues of misuse
	S				Expected to explain the internal rules used for generation	

4.4. Discussion: Current state, challenges, and prospects for future research

In terms of AI capabilities, capabilities like recognition, prediction, and generation have garnered intensive research interest and achieved significant technical enhancements. Reasoning is often needed implicitly when AI performs other tasks. As a separate AI capability, reasoning is the least studied and developed. However, the advent of ChatGPT and other LLMs greatly facilitates the development of reasoning AI (Angel et al., 2023). We anticipate to see the rapid growth of AI-powered reasoning with

https://doi.org/10.1017/pds.2024.227 Published online by Cambridge University Press

the continuous development of AGI, enhancing the development of HAIC. Particularly, cross-modal reasoning is expected to enhance cross-modal design generation, which may allow humans to ideate at an abstract level and AI to create designs accordingly at a concrete level, accelerating design generation and broadening design exploration.

In recent years, AI's interactive attributes have seen notable advancements, particularly in sensing and predictability (Endsley, 2023), adaptability (Wang et al., 2023), and awareness sharing (Jiang et al., 2023). However, compared to the task-oriented capabilities, AI's interactive attributes are still less developed (Song and Zurita et al., 2022). In the future, more research efforts should be invested to enhance context-awareness and agility of AI when faced with evolving design scenarios. At the core of AI's interactivity lies the capability to detect and model problem and human conditions. Although we argue that the other two dimensions of AI, the Intelligence Scope and Cognition Mode, have limited influence on the necessity of the interactive attributes, they may affect their implementation.

Moreover, significant progress has been made in the design and implementation of trust enablers for HAIC, which also present distinct challenges. Researchers have explored multiple perspectives to design and improve AI transparency (Vössing et al., 2022), empathy (Srinivasan and González, 2022), reliability (Inel, 2023), interpretability (Ross et al., 2021), and ethicality (Schelble et al., 2022), improving human trust and satisfaction towards AI (Schelble et al., 2022). However, obstacles have also been identified. For example, a recent study argued that since empathy is rooted in distinctive human experience, implementing AI empathy can be intricate and needs to be approached carefully, as it has the potential to backfire if the design of empathy cannot generalize well across different groups of people (Shao, 2023). More future efforts should be made to understand, design, and implement trust enablers for trustworthy AI development to boost HAIC.

The integration of advanced task capabilities, interactive attributes, and trust enablers can endow AI with "human traits" in HAIC. This offers several advantages: (1) it promotes more natural and engaging interactions by reflecting human conversational patterns, (2) it eases the integration of teams by making AI's roles and contributions more relatable and understandable to humans, (3) it increases role adaptability as AI can modify its role responsively, showing human-like initiative and flexibility, and (4) it fosters a shared understanding since AI can both interpret human feedback and be interpreted more effectively. Moreover, many AI agents fall short due to their lack of empathy, limiting their capacity to grasp cultural contexts and the personalities of humans, such as users or team members. AI empathy may be enhanced by incorporating psychologists and anthropologists into the development teams for human-centred design. To successfully create multidisciplinary hybrid teams, collecting insights from various organizational managers through interviews or surveys can offer crucial perspectives in defining the roles of AI.

Compared to the AI frameworks introduced in previous studies (Dellermann et al., 2019; Dubey et al., 2020; Seeber et al., 2020), our proposed framework for HAIC provides a multi-perspective guideline for outlining the expected capabilities, interactive attributes, and trust enablers of AI, as detailed in the preceding section. However, our current framework has several limitations. Firstly, it delineates the expected AI capabilities under various HAIC settings from multiple perspectives, yet it lacks detailed instructions for endowing AI with these capabilities. Secondly, the framework addresses only three aspects of task characteristics and human-AI interaction, leaving out others such as task allocation and information flow illustrated in Figure 1. Thirdly, the framework is constructed from the perspective of AI design, without considering the design of human capabilities and team structures for different HAIC settings. Future research should aim to refine the proposed AI framework's implementation strategies and expand it by integrating the overlooked elements and perspectives.

5. Conclusion

In conclusion, this paper has presented a nuanced framework for the classification and design of AI to foster human-AI collaboration (HAIC), offering a significant contribution to the field. We have delineated a new scheme for AI role classification that is both unified and comprehensive, capable of distinctly categorizing any AI use case across a spectrum of abilities. Building upon this classification, we proposed a detailed framework for AI design, highlighting the expected capabilties, interactive attributes, and trust enablers of AI in various HAIC contexts. This paper underscores the importance of designing AI systems that are capable of working alongside humans in a collaborative, supportive, and

https://doi.org/10.1017/pds.2024.227 Published online by Cambridge University Press

trustworthy manner. While the transformative potential of such AI is vast, we also recognize the limitations and challenges that lie ahead. Future research should focus on refining the proposed framework to include more granular instructions for AI capability development, as well as expanding the framework to encompass additional aspects of human-AI interaction and the design of human roles and team structures. The path forward for AI is not just in technological advancement, but in fostering synergistic partnerships with humans, leveraging the strengths of both to achieve unprecedented levels of innovation and efficiency.

References

Bittner, E., Oeste-Reiß, S. and Leimeister, J.M., 2019. Where is the bot in our team? Toward a taxonomy of design option combinations for conversational agents in collaborative work.

Bruemmer, D.J., Marble, J.L. and Dudenhoeffer, D.D., 2002, September. Mutual initiative in human-machine teams. In Proceedings of the IEEE 7th conference on human factors and power plants (pp. 7-7). IEEE.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. arXiv preprint arXiv:2303.12712.

Caldwell, S., Sweetser, P., O'Donnell, N., Knight, M.J., Aitchison, M., Gedeon, T., Johnson, D., Brereton, M., Gallagher, M. and Conroy, D., 2022. An agile new research framework for hybrid human-AI teaming: Trust, transparency, and transferability. ACM Transactions on Interactive Intelligent Systems (TiiS), 12(3), pp.1-36.

Callaway E. 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. Nature. 2020 Dec;588(7837):203-204. https://dx.doi.org/10.1038/d41586-020-03348-4. PMID: 33257889.

Chong, L., Raina, A., Goucher-Lambert, K., Kotovsky, K. and Cagan, J., 2023. The evolution and impact of human confidence in artificial intelligence and in themselves on ai-assisted decision-making in design. Journal of Mechanical Design, 145(3), p.031401.

Chong, L., Zhang, G., Goucher-Lambert, K., Kotovsky, K. and Cagan, J., 2022. Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of AI advice. Computers in Human Behavior, 127, p.107018.

Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S. and Ebel, P., 2019. The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems. Hawaii International Conference on System Sciences (HICSS).

Deloitte. (2020). Insights 2020. The social enterprise in a world disrupted: Leading the shift from survive to thrive, 2021 DELOITTE GLOBAL HUMAN CAPITAL TRENDS, 64 pages, 9th of December.

Dubey, A., Abhinav, K., Jain, S., Arora, V. and Puttaveerana, A., 2020, February. HACO: a framework for developing human-AI teaming. 13th Innovations in Software Engineering Conference (pp. 1-9).

Endsley, M.R., 2023. Supporting Human-AI Teams: Transparency, explainability, and situation awareness. Computers in Human Behavior, 140, p.107574.

Filippi, S., 2023. Measuring the impact of ChatGPT on fostering concept generation in innovative product design. Electronics, 12(16), p.3535.

Gyory, J.T., Song, B., Cagan, J. and McComb, C., 2021. Communication in AI-assisted teams during an interdisciplinary drone design problem. Proceedings of the Design Society, 1, pp.651-660.

Inel, O., Draws, T. and Aroyo, L., 2023, November. Collect, measure, repeat: Reliability factors for responsible AI data collection. AAAI Conf. on Human Computation and Crowdsourcing (Vol. 11, No. 1, pp. 51-64).

Jiang, J., Karran, A.J., Coursaris, C.K., Léger, P.M. and Beringer, J., 2023. A situation awareness perspective on human-AI interaction: Tensions and opportunities. International Journal of Human–Computer Interaction, 39(9), pp.1789-1806.

Liu, M., Hu, Y. (2023). Application Potential of Stable Diffusion in Different Stages of Industrial Design. Artificial Intelligence in HCI. Lecture Notes in Computer Science, vol 14050. Springer, Cham.

Luo, J., 2022. Data-driven innovation: What is it?. IEEE Transactions on Engineering Management, 70(2), pp.784-790.

Luo, J., 2023. Designing the future of the fourth industrial revolution. Journal of Engineering Design, 34(10), 779-785.

Luo, J., Sarica, S., & Wood, K. L. (2021). Guiding data-driven design ideation by knowledge distance. Knowledge-Based Systems, 218, 106873, 2021.

McDermotta, P.L., Walkera, K.E., Dominguez, C.O., Nelsonb, A. and Kasdaglis, N., 2017. Quenching the thirst for human-machine teaming guidance: Helping military systems acquisition leverage cognitive engineering research. In 13th International Conference on Naturalistic Decision Making (pp. 236-240).

Memmert, L. and Bittner, E., 2022. Complex Problem Solving through Human-AI Collaboration: Literature Review on Research Contexts.

Nichol, A., Jun, H., Dhariwal, P., Mishkin, P. and Chen, M., 2022. Point-e: A system for generating 3d point clouds from complex prompts. arXiv preprint arXiv:2212.08751.

OpenAI., 2023. ChatGPT (Mar 14 version) [Large language model]. https://chat.openai.com/chat.

https://doi.org/10.1017/pds.2024.227 Published online by Cambridge University Press

Papachristos, E., Skov Johansen, P., Møberg Jacobsen, R., Bjørn Leer Bysted, L. and Skov, M.B., 2021. How do People Perceive the Role of AI in Human-AI Collaboration to Solve Everyday Tasks?. 1st International Conference of the ACM Greek SIGCHI Chapter (pp. 1-6).

Peeters, M.M., van Diggelen, J., Van Den Bosch, K., Bronkhorst, A., Neerincx, M.A., Schraagen, J.M. and Raaijmakers, S., 2021. Hybrid collective intelligence in a human–AI society. AI & society, 36, pp.217-238.

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M. and Sutskever, I., 2021, July. Zero-shot text-to-image generation. In International Conference on Machine Learning (pp. 8821-8831). PMLR.

Regenwetter, L., Nobari, A. H., & Ahmed, F. (2022). Deep generative models in engineering design: A review. Journal of Mechanical Design, 144(7), 071704.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B., 2022. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10684-10695).

Ross, A., Chen, N., Hang, E.Z., Glassman, E.L. and Doshi-Velez, F., 2021, May. Evaluating the interpretability of generative models by interactive reconstruction. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-15).

Schelble, B.G., Lopez, J., Textor, C., Zhang, R., McNeese, N.J., Pak, R. and Freeman, G., 2022. Towards ethical AI: Empirically investigating dimensions of AI ethics, trust repair, and performance in human-AI teaming. Human Factors, p.00187208221116952.

Seeber, I., Bittner, E., Briggs, R.O., De Vreede, T., De Vreede, G.J., Elkins, A., Maier, R., Merz, A.B., Oeste-Reiß, S., Randrup, N. and Schwabe, G., 2020. Machines as teammates: A research agenda on AI in team collaboration. Information & management, 57(2), p.103174.

Shao, R., 2023, April. An Empathetic AI for Mental Health Intervention: Conceptualizing and Examining Artificial Empathy. In Proceedings of the 2nd Empathy-Centric Design Workshop (pp. 1-6).

Siddharth, L., Blessing, L., & Luo, J. (2022). Natural language processing in-and-for design research. Design Science, 8, e21.

Siemon, D., 2022. Elaborating team roles for artificial intelligence-based teammates in human-AI collaboration. Group Decision and Negotiation, 31(5), pp.871-912.

Song, B., Yuan, C., Permenter, F., Arechiga, N. and Ahmed, F., 2023. Surrogate Modeling of Car Drag Coefficient with Depth and Normal Renderings. arXiv preprint arXiv:2306.06110.

Song, B., Gyory, J.T., Zhang, G., Zurita, N.F.S., Stump, G., Martin, J., Miller, S., Balon, C., Yukish, M., McComb, C. and Cagan, J., 2022. Decoding the agility of artificial intelligence-assisted human design teams. Design Studies, 79, p.101094.

Song, B., Soria Zurita, N.F., Nolte, H., Singh, H., Cagan, J. and McComb, C., 2022. When faced with increasing complexity: the effectiveness of artificial intelligence assistance for drone design. Journal of Mechanical Design, 144(2), p.021701.

Song, B., Soria Zurita, N.F., Nolte, H., Singh, H., Cagan, J. and McComb, C., 2021. Addressing challenges to problem complexity: Effectiveness of AI assistance during the design process. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.

Srinivasan, R. and González, B.S.M., 2022. The role of empathy for artificial intelligence accountability. Journal of Responsible Technology, 9, p.100021.

Tan, S.C., Chen, W. and Chua, B.L., 2023. Leveraging generative artificial intelligence based on large language models for collaborative learning. Learning: Research and Practice, pp.1-10.

Turing, A.M., 2012. Computing machinery and intelligence (1950). The Essential Turing: the Ideas That Gave Birth to the Computer Age, pp.433-464.

Viros-i-Martin, A. and Selva, D., 2019. Daphne: A virtual assistant for designing earth observation distributed spacecraft missions. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, pp.30-48.

Vorobeva, D., Costa Pinto, D., António, N. and Mattila, A.S., 2023. The augmentation effect of artificial intelligence: can AI framing shape customer acceptance of AI-based services? Tourism, pp.1-21.

Wang, L., Zhang, X., Li, Q., Zhang, M., Su, H., Zhu, J. and Zhong, Y., 2023. Incorporating neuro-inspired adaptability for continual learning in artificial intelligence. Nature Machine Intelligence, pp.1-13.

Zhang, G., Raina, A., Cagan, J. and McComb, C., 2021. A cautionary tale about the impact of AI on human design teams. Design Studies, 72, p.100990.

Zhu, Q., & Luo, J. (2023a). Generative transformers for design concept generation. Journal of Computing and Information Science in Engineering, 23(4), 041003.

Zhu, Q., & Luo, J. (2023b). Toward Artificial Empathy for Human-Centered Design: A Framework. Journal of Mechanical Design, 146(6), 061401.

Zhu, Q., Zhang, X., & Luo, J. (2023). Biologically Inspired Design Concept Generation Using Generative Pre-Trained Transformers. Journal of Mechanical Design, 145(4), 041409.

https://doi.org/10.1017/pds.2024.227 Published online by Cambridge University Press