Engineering and Technology Journal e-ISSN: 2456-3358
Volume 09 Issue 11 November-2024, Page No.- 5549-5557
DOI: 10.47191/etj/v9i11.14, I.F. – 8.227

Designing and Evaluating a Metacognitive AI System for Enhanced Human-AI Collaboration: A Woz Approach

Ayse Kok Arslan
Researcher, Silcion Valley Chapter- Oxford ALumni

ABTSRACT: This study leverages recent theoretical and methodological advancements to design and evaluate a Metacognitive Artificial Intelligence (MAI) system. This study employs the WOz approach from the field of human-computer interaction, to explore the design of technological systems that support human-AI collaboration. This study supports the view that human-AI collaboration in research with multidisciplinary joint forces will facilitate empirical evidence and design work to articulate human-AI collaboration.

INTRODUCTION

The widespread adoption of Artificial Intelligence (AI) across diverse domains of work and daily life highlights the critical need to equip individuals with the skills to collaborate effectively not only with their peers but also within socio-technological ecosystems that integrate AI. Harnessing the complementary strengths of humans and AI offers the potential to generate unique insights and advance collaborative practices.

This research builds upon contemporary theoretical and methodological progress to conceptualize and assess a Metacognitive Artificial Intelligence (MAI) framework. Following a concise review of intelligent systems, the study introduces the Human-AI Shared Regulation in Learning (HASRL) model, which serves as a foundational paradigm for examining and enhancing human-AI interactions.

Utilizing the Wizard of Oz (WOz) methodology from the domain of human-computer interaction, this study investigates the development of technological systems designed to facilitate human-AI collaboration. The findings underscore the significance of interdisciplinary efforts in advancing empirical research and design initiatives to articulate and improve the dynamics of human-AI cooperation.

REVIEW OF EXISTING WORK

The rapid integration of AI into various dimensions of daily life necessitates a fundamental reassessment and refinement of existing theoretical frameworks to effectively support the synergy between human and artificial intelligence in collaborative processes (Gašević et al., 2023; Nguyen et al., 2024).

Human-AI collaboration can be understood as an agentic process, wherein individuals actively and strategically manage their interactions with AI systems and tools (Zimmerman, 1989). While the transformative potential of multimodal data and AI-driven methodologies in fostering human-AI collaboration is widely acknowledged, the design and implementation of systems that harness these capabilities remain limited.

Creating a robust AI-enhanced system that facilitates effective human-AI collaboration poses significant challenges. These challenges arise from the intricate task of developing systems that are not only grounded in solid theoretical principles but also demonstrate advanced technological capabilities to address the complex dynamics of collaboration.

Addressing such complexities calls for the integration of hybrid intelligence—a fusion of human and artificial intelligence designed to amplify human cognitive capabilities rather than replace them with AI systems. By combining human intentionality and expertise with machine intelligence, hybrid intelligence strives to create seamless cooperation between AI agents and human users (Akata et al., 2020).

Holstein et al. (2020) propose a comprehensive framework of dimensions to characterize hybrid human-AI adaptations, offering valuable insights into how these approaches can inspire innovative possibilities.

Before progressing further into the exploration of intelligent systems, it is essential to establish a clear definition of intelligence as a foundational concept.

REVIEW OF EXISTING THEORIES ON INTELLIGENCE SYSTEMS

Minsky's 1968 definition of artificial intelligence offers a foundational view of the mind as a collection of static, vertical programs collectively enabling "intelligence": "AI is the science of making machines capable of performing tasks that would require intelligence if done by human beings." This concept aligns with Turing's earlier vision, proposed in his seminal 1950 paper, which suggested that machines could

acquire skills through a learning process analogous to that of human children.

With the evolution of AI research, the philosophy of intelligence has shifted significantly. The resurgence of machine learning in the 1980s, its growing intellectual prominence in the early 2000s, and its dominance by the late 2010s through Deep Learning positioned the connectionist-inspired Tabula Rasa as a prevailing philosophical framework. This approach draws on Locke's concept of the mind as a blank slate—an adaptable, general process capable of transforming experience into behavior, knowledge, and skills—shaping the history of cognitive science and modern AI paradigms.

In 2007, Legg and Hutter synthesized over 70 definitions of intelligence into a concise formulation: "Intelligence measures an agent's ability to achieve goals in a wide range of environments." They observed that no comprehensive survey of intelligence definitions or evaluation methods had been published at the time. A decade later, Hernández-Orallo addressed this gap with an extensive survey and a detailed book on AI evaluation. These works highlighted two fundamental characteristics often embedded in intelligence definitions: **task-specific proficiency** ("achieving goals") and **general adaptability** ("in a wide range of environments"). An intelligent agent, according to this framework, demonstrates high proficiency across diverse tasks, including the ability to acquire new skills for previously unknown challenges, embodying true generality.

The structure of human intelligence, as outlined in major theories, follows a hierarchical model with three interconnected layers:

- **General intelligence (g factor)** at the top, representing extreme generalization.
- **Broad abilities** in the middle, encompassing domain-specific generalization.
- **Specialized skills** or task-specific abilities at the bottom, representing local generalization or cases with no generalization.

In this hierarchical framework, the g factor encapsulates the highest level of generalization, while broad cognitive abilities and specialized skills reflect varying degrees of adaptability across domains and tasks. For instance, achieving proficiency across a diverse set of video games serves as a practical example of broad generalization, with skill acquisition for novel tasks demonstrating the agent's potential for extreme generalization.

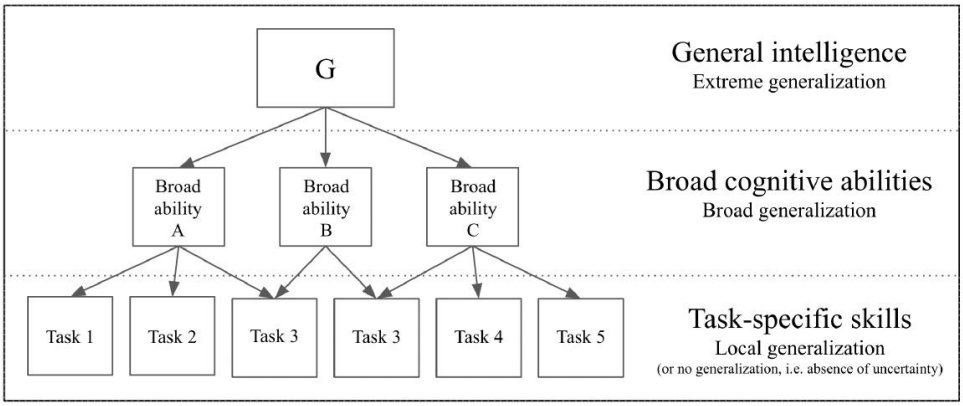

Figure 1. Hierarchical model of human intelligence

"Generalization" or "generalization power" in the context of AI refers to the ability of an AI system to handle situations or tasks that differ from those it has previously encountered. Since the concept of "previously encountered situations" can be ambiguous, it is useful to categorize generalization into five distinct types, each addressing different scenarios and levels of adaptability:

1. System-Centric Generalization

This pertains to a learning system's capacity to handle situations it has not directly encountered during training. It aligns with the formal notion of generalization error in statistical learning theory. For example, a machine learning classifier trained on a dataset of NNN samples demonstrates system-centric generalization if it accurately classifies images outside the training set.

2. Developer-Aware Generalization

This represents the ability of a system to handle situations unfamiliar to both the system itself and its developer. Here, the system goes beyond merely adapting to unseen data by addressing scenarios not envisioned by its creator.

3. Local Generalization (or "Robustness")

Local generalization reflects a system's capacity to adapt to new inputs within a familiar distribution for a specific task. It requires the system to tolerate anticipated perturbations or variations in a fixed context, provided the data samples sufficiently represent the known distribution.

4. Broad Generalization (or "Flexibility")

Broad generalization refers to the system's ability to handle a diverse category of tasks and environments without additional human intervention. This includes unforeseen

situations, showcasing a system's flexibility to adapt to broad, previously undefined scenarios.

5. Extreme Generalization

Extreme generalization describes open-ended systems capable of addressing entirely new tasks that share only abstract similarities with previously encountered situations. Such systems exhibit the highest level of adaptability, transcending specific domains or contexts to operate in a wide scope of tasks.

The progression from system-centric to extreme generalization reflects an increasing level of adaptability and complexity, where higher levels address broader, less predictable scenarios.

In this framework, Solomonoff et al. (2024) propose that generalization operates on the assumption of:

- **A fixed universal Turing machine**: All relevant programs—skill programs, task-specific programs, and components of the intelligent system—are executed on this machine.
- **A fixed "situation space" and "response space"**: These spaces represent the input scenarios the system may encounter and the potential responses it can generate.

This theoretical model underpins the design and evaluation of generalization in AI systems, establishing a structured foundation to explore their potential to learn, adapt, and handle new challenges.

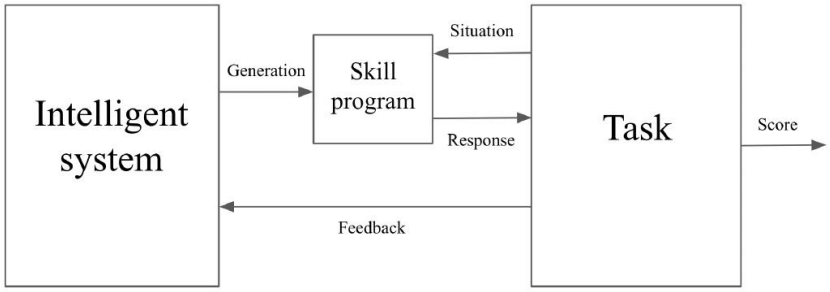

Figure 2. Overview of an intelligent system

Solomonoff et al. (2024) present several critical observations regarding their conceptualization of the mind, emphasizing the relationship between intelligence, generalization, and adaptability in the context of high-intelligence systems. These insights are particularly relevant when considering the shift toward hybrid intelligence systems that combine human and AI capabilities.

Key observations by Solomonoff et al. (2024) include, but are not limited to:

1. **Efficiency and Generalization in High-Intelligence Systems**
   A high-intelligence system is defined by its ability to generate high-skill solution programs for tasks with high generalization difficulty (e.g., tasks with significant uncertainty about the future). This requires the system to make highly efficient use of its available information and priors, enabling it to address unknown parts of the situation space with minimal experience.
2. **Scope-Dependent Measurement of Intelligence**
   Intelligence is inherently tied to the choice of task scope and the value function assigned to these tasks. It may optionally include a sufficient skill level requirement for tasks within this scope. The adaptability and relevance of an intelligent system are determined by the breadth and depth of its defined task space.
3. **Adaptation as a Core Component of Intelligence**
   Intelligence necessitates adaptation—extracting and operationalizing information from past experiences to navigate future uncertainties. A system with pre-existing proficiency for evaluation tasks has low generalization difficulty and, thus, scores poorly on intelligence metrics emphasizing adaptability and learning.
4. **Beyond Curve-Fitting**
   Intelligence transcends mere curve-fitting. A system that generates the simplest skill program consistent with known data can only handle tasks with zero generalization difficulty. To address future uncertainties, an intelligent system must produce behavioral programs capable of adapting to new, unforeseen situations.

Building upon Solomonoff et al.'s observations, advancing AI-human collaboration necessitates a rethinking of how AI systems are designed and utilized. Hybrid intelligence—blending human expertise with machine intelligence—requires researchers to push beyond conventional boundaries to create systems that achieve synergy between humans and AI.

Recent developments in AI have introduced innovative ways to utilize multimodal data from diverse sources, such as:

- Digital traces like log data (Cho & Yoo, 2017)

- Physiological measures, including eye gazes (Taub & Azevedo, 2016)
- AI-driven multimodal analytics, offering scalable data processing, analysis, and predictive capabilities (Nguyen, Järvelä, Rosé, et al., 2022; Sharma & Giannakos, 2020).

These advancements allow for sophisticated insights and predictive modeling but also reveal significant gaps. Existing theoretical frameworks were developed before such multimodal capabilities emerged, and they fall short in addressing the complexities of AI-human collaboration in research and application contexts.

While there has been progress in using AI to explore and support user self-regulation (Fan et al., 2022), little has been done to understand how researchers can effectively collaborate with AI to integrate human and machine intelligence. Filling this gap is essential for advancing hybrid intelligence and realizing its potential for transformative impacts in both theoretical and applied domains.

This study seeks to address these challenges by establishing new approaches to integrating AI into collaborative processes, fostering a deeper understanding of AI-human synergy, and shaping the evolution of hybrid intelligence systems.

THEORETICAL FRAMEWORK

The integration of AI within socio-technical systems has inspired a variety of conceptual frameworks aimed at leveraging AI affordances and multimodal data to examine and address complex processes. These frameworks build on theoretical underpinnings of socio-technical systems, such as Leavitt's (1964) model, and provide mechanisms to study and enhance human-AI collaboration. Below is a synthesis of these theoretical perspectives and their relevance to advancing AI-enabled collaborative systems.

Leavitt's (1964) socio-technical systems model provides a foundational lens for understanding the interplay between people, technology, tasks, and structure, all of which dynamically interact to produce emergent phenomena (Truex & Baskerville, 1998).

- People: Users, developers, providers, and regulators shape and interact with AI systems, influencing their design and usage.
- Technology: Includes multimodal tools such as affective computing techniques, chatbots, sentiment analysis, and avatars, which serve as problem-solving enablers.
- Task: Encompasses the practices and sequences required to achieve overarching goals, which are sensitive to contextual and situational demands.
- Structure: Represents cultural, ethical, and institutional principles that shape governance, communication, and authority within AI systems.

These components are deeply interconnected, producing emergent properties, such as consciousness in neural networks or collective behaviors in natural systems (Gloor, 2006). The interplay between technology and structure, in particular, influences the ethical, legal, and societal implications of AI integration, while the interaction between tasks and structure defines how AI aligns with societal values and organizational goals.

Within this context, the frameworks for collaboration and regulation in socio-technical AI systems can be summarized as follows:

1. Collaboration as Cyclical and Recursive Processes

Molenaar et al. (2021) emphasize that all types of collaboration involve cyclical and recursive processes. Regulatory responses emerge dynamically, shaped by the timing and sequence of actions observed during collaborative work. This perspective underscores the importance of identifying and responding to collaboration breakdowns or disruptions in real-time.

2. Winne and Hadwin's Four-Phase Model

Winne and Hadwin's (1998) model identifies four loosely sequenced and recursive phases of self-regulation:

- Task definition,
- Goal setting,
- Strategic planning, and
- Adaptation based on monitoring and evaluation.

This framework, as expanded by Järvelä et al. (2018), highlights the role of collective monitoring and evaluation in guiding group decision-making and adapting collaborative practices in response to disruptions. This makes it particularly relevant for designing human-AI systems where feedback loops are critical for ensuring effective collaboration.

3. Trigger Regulation Framework

Järvelä and Hadwin (2024) introduce the concept of trigger events, defined as situations or disruptions that may inhibit collaboration and require regulatory responses. This framework facilitates examining the timing, frequency, and sequences of regulatory traces (Saint et al., 2020) to identify struggling individuals or teams and provide timely interventions or prompts to optimize collaboration.

In this framework, trigger events are central to understanding how collaborative systems, particularly AI-supported ones, adapt to situational disruptions.

4. Shared Regulatory Processes for Human-AI Systems

Järvelä et al. (2023) expand on the trigger regulation framework to propose a structured model for understanding shared regulatory processes between humans and AI. Their framework emphasizes:

- Mechanisms of interaction between human and AI regulatory processes,
- Responsiveness to trigger events, and
- Tuning AI models to enhance human regulatory capacities.

Inspired by these frameworks, this study adopts the Wizard of Oz (WOz) methodology from human-computer interaction to explore the design of socio-technological systems that support human-AI collaboration. The WOz approach simulates AI behavior by having a human operator perform AI tasks behind the scenes, allowing researchers to investigate:

- How users interact with AI systems,
- The efficacy of proposed designs, and
- The dynamics of regulatory processes in hybrid collaboration settings.

This approach enables researchers to refine AI system designs iteratively, ensuring alignment with the theoretical principles outlined in the frameworks above.

RESEARCH METHODOLOGY

The Echeloned Design Science Research (eDSR) framework, introduced by Tuunanen et al. (2024), is a groundbreaking evolution of the traditional DSR methodology. It specifically addresses the challenges of complexity, time, and resource intensity associated with traditional DSR, particularly in the context of developing AI-enhanced systems for human collaboration.

By breaking down the research process into hierarchical phases, or "echelons," eDSR enables researchers to focus on specific facets of a project in a logical and modular manner. This segmentation enhances clarity, improves manageability, and minimizes risk.

The five echelons of eDSR include;

1. **Problem Analysis Echelon**:
    - Identifies and refines the core research problem.
    - Establishes a foundation by aligning the problem with stakeholder needs and socio-technical considerations.
2. **Objectives and Requirements Definition Echelon**:
    - Outlines specific goals and requirements for the artefact.
    - Ensures alignment with theoretical frameworks, such as human-AI shared regulation or trigger-based collaboration models.
3. **Design and Development Echelon**:
    - Focuses on iterative artefact creation, leveraging user-centered design principles and prototyping.
    - Iterations allow for feedback integration, ensuring the artefact evolves in alignment with its objectives.
4. **Demonstration Echelon**:
    - Showcases the artefact in simulated or real-world environments to validate its core functionality and feasibility.
    - May involve pilot studies or Wizard of Oz (WOz) experiments to simulate AI responses and user interactions.
5. **Evaluation Echelon**:
    - Conducts rigorous assessments of the artefact's utility, usability, and effectiveness.
    - Employs both qualitative and quantitative methods to validate its impact on collaboration and regulatory processes.
    -

The eDSR framework is particularly suited for research exploring human-AI shared regulation, as it:

- Facilitates the integration of multimodal data (e.g., behavioral traces, interaction logs) into iterative design processes.
- Aligns with frameworks like trigger event concepts and Winne & Hadwin's model by incorporating their insights into the problem analysis and design phases.
- Enables the design of adaptive AI tools that can dynamically respond to trigger events, thereby enhancing collaborative processes.

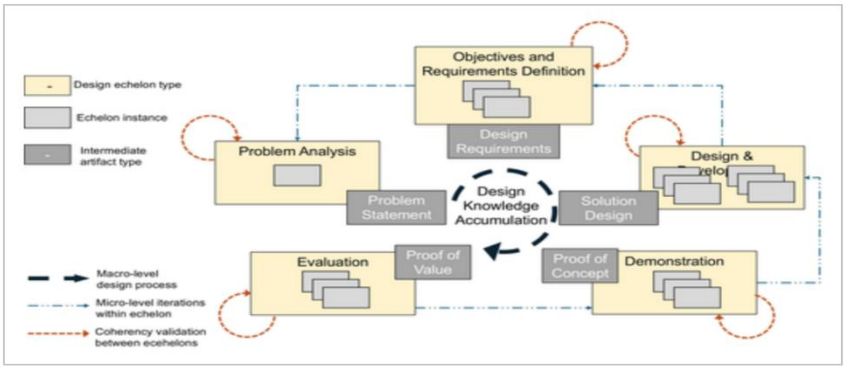

Figure 3. Overview of eDSR Methodological Framework

The integration of the Wizard of Oz (WOz) paradigm into the Echeloned Design Science Research (eDSR) framework enhances its ability to address the complexities and resource demands of developing sophisticated AI systems. By

adopting a **"fail early, fail often"** approach, eDSR enables researchers to mitigate risks early in the research process while maintaining adaptability to diverse project needs. This combination of structure and flexibility ensures that iterative feedback loops remain central, keeping research aligned with stakeholder needs and real-world applications.

Each echelon in the eDSR framework incorporates dedicated validation points, encouraging researchers to test assumptions and design decisions at every stage. By validating artefact objectives, user needs, and system functionality incrementally, the framework reduces the likelihood of costly errors later in the process. The inclusion of these checkpoints enables:

- **Early Problem Identification**: Issues are addressed during initial design iterations rather than after significant investments in development.
- **Iterative Refinement**: Artefacts can evolve through successive cycles of testing, feedback, and redesign.
- **Continuous Stakeholder Engagement**: Regular checkpoints ensure user and stakeholder involvement throughout the process, fostering trust and alignment.

Wizard of Oz (WOz) Paradigm in eDSR

The WOz paradigm plays a critical role in advancing the eDSR methodology, particularly within the Design and Development and Demonstration Echelons. By simulating AI system functionality through human intervention, WOz allows researchers to study interactions and collect valuable feedback without requiring a fully functional system.

An example of WOz Workflow in eDSR can be provided as follows:

1. **Problem Analysis Echelon**:
    - Identify interaction scenarios that would benefit from AI support (e.g., shared regulation in collaborative settings).
2. **Objectives and Requirements Echelon**:
    - Define the system's expected functionalities and user interaction workflows.
3. **Design and Development Echelon**:
    - Use WOz to prototype AI behavior (e.g., a simulated AI assistant that responds to user queries).
4. **Demonstration Echelon**:
    - Conduct experiments where participants interact with the simulated system. For instance, a researcher ("wizard") may type responses to simulate natural language understanding and generation.
5. **Evaluation Echelon**:
    - Analyze user feedback and interaction data to assess the artefact's effectiveness and inform further refinements.

By incorporating the WOz paradigm into the eDSR methodology, researchers can significantly streamline the development of AI-enhanced systems. The ability to test and iterate on designs early and efficiently fosters a deeper understanding of user needs, reduces risks, and ensures that the final system is both functional and effective. Together, eDSR and WOz provide a robust framework for tackling the complexities of socio-technical systems while maintaining empirical rigor and practical relevance.

DESIGN REQUIREMENTS AND PRINCIPLES FOR METACOGNITIVE ARTIFICIAL INTELLIGENCE (MAI)

The research aims to design a **Metacognitive AI (MAI)** system that can enhance collaborative learning and shared regulation processes among users. To achieve this, the design of the system needs to align with established principles that guide effective collaboration and metacognitive engagement. Building on the work of Järvelä et al. (2015), the **initial design requirements** focus on three critical principles:

1. **Awareness**:
    - **Goal**: To increase users' awareness of their own and others' collaboration processes. This principle fosters metacognitive engagement, allowing users to monitor, reflect on, and assess their collaborative strategies and progress. The ability to recognize how one is contributing to the collaboration (or where they may be falling short) is central to effective teamwork and learning.
2. **Externalization**:
    - **Goal**: To support the externalization of users' collaboration processes on a social plane. This principle enables users to share their thoughts, strategies, and approaches openly, thus providing greater insight into each participant's reasoning and decision-making. Such externalization is essential for collaborative knowledge co-construction and refining skills.
3. **Prompting Regulation**:
    - **Goal**: To prompt the acquisition and activation of regulatory processes, guiding users toward more effective collaboration. This principle focuses on the MAI system's ability to support self-regulation by providing cues or feedback that prompt users to adjust their behavior when necessary, thereby improving the overall collaboration process.

These principles together set the foundation for developing an AI system that not only supports the task at hand but also

enhances the metacognitive aspects of collaboration, enabling users to be more conscious of and engaged in their interactions.

The **Wizards of Oz (WOz) prototype** serves as an initial demonstration of how the proposed MAI system could interact with users and support the collaboration process. The WOz prototype simulates an autonomous, proactive speech agent embedded within an iPad placed on the user's table, acting as if it can engage in real-time conversations with users. The key aspects of this prototype design are as follows:

1. **Proactive Interaction**:
    - Unlike traditional systems that only respond to user input, the MAI agent in this prototype is designed to **initiate** interactions proactively. This behavior simulates an advanced, autonomous system capable of determining when and how to interact based on context or observed patterns in the conversation.
2. **Speech-based Interface**:
    - The prototype leverages **speech-based interaction** to facilitate natural communication between users and the agent. This mirrors how an advanced AI system would operate in real-world collaborative settings, enabling the MAI to seamlessly integrate into discussions without disrupting the flow.
3. **Simulating Autonomous Actions**:
    - In this setup, a **human "wizard"** operates behind the scenes, controlling the system's responses without the user's knowledge. The wizard's goal is to simulate the agent's behavior based on the predetermined principles (awareness, externalization, and prompting regulation). The wizard listens to user conversations and provides the illusion that the agent is interacting intelligently.
4. **Real-Time Feedback and Adaptation**:
    - As the prototype operates in real-time, users interact with the speech agent, providing valuable insights into how well the system's behaviors align with the design principles. For example, the MAI system could prompt users to reflect on their contribution to the collaboration, or provide feedback on how they could improve their regulatory processes.

Prototype Evaluation and Feedback

The **WOz prototype** provides a controlled environment to explore how the MAI system might perform in a real-world scenario, allowing researchers to collect feedback from participants without needing to develop a fully functional AI system at the outset. During the demonstration:

- **User Interaction**: Participants engage with the prototype, interacting with the speech agent as though it were fully autonomous.
- **Observational Data**: Researchers observe how participants respond to the proactive interactions initiated by the MAI system and assess whether these interactions align with the principles of awareness, externalization, and prompting regulation.
- **User Feedback**: After interacting with the prototype, participants provide insights into their experience, helping refine design requirements and system behavior for future iterations.

The integration of the **Wizards of Oz (WOz)** paradigm within the **eDSR framework** for developing the **Metacognitive AI (MAI)** system enhances the ability to explore, prototype, and refine AI-driven collaborative tools without the need for extensive technological investments upfront. By simulating proactive interaction behaviors and evaluating real-time feedback from participants, this approach allows researchers to verify the feasibility and functionality of the proposed solution, ensuring that the final system is both practical and aligned with the research objectives.

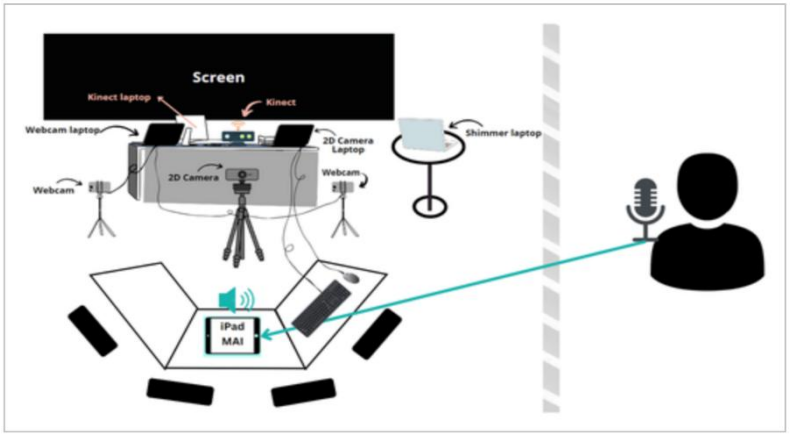

Figure 4. Overview of Wizards of Oz Prototype

DISCUSSION AND SUGGESTIONS

This interdisciplinary study explores the necessary methodological framework for the development of a Metacognitive Agent (MAI) designed to support collaborative processes in face-to-face environments. The research adopts the eDSR methodology (Tuunanen et al., 2024), emphasizing the iterative refinement of design principles for MAI through the use of a Wizards of Oz (WOz) prototype. This methodological approach enables the exploration of MAI's role in enhancing collaboration and metacognition, critical aspects of complex human learning and interaction.

Hybrid intelligence enables humans to learn from and reinforce each other, creating an adaptive feedback loop that promotes mutual growth and understanding (Akata et al., 2020).

This research aims to bridge theoretical insights with practical design strategies, contributing to the development of AI technologies that can support complex cognitive and metacognitive activities in collaborative contexts. By refining the design principles for MAI through an iterative process facilitated by the eDSR methodology and the WOz prototype, the study endeavors to create a more adaptive, flexible, and user-centric system that not only augments collaboration but also promotes deeper metacognitive engagement.

The integration of hybrid intelligence, with AI acting as a complementary tool for human collaborators, holds significant promise for reshaping the landscape of collective learning and problem-solving. This work aims to inform the design of AI systems that not only perform tasks but also enhance human cognitive and socioemotional capabilities, ultimately supporting more effective, fulfilling, and adaptive collaborative experiences.

FINAL REMARKS AND FUTURE DIRECTIONS

Interdisciplinary collaborations that integrate the strengths of both humans and AI hold immense potential to transform the field of research and significantly advance our understanding of collaborative processes. By combining the cognitive, emotional, and social capabilities of humans with the computational power and analytical precision of AI, researchers and technologists can develop tools that not only optimize collaboration but also ensure that these tools are ethical, equitable, and effective for diverse user groups.

The study argues that the fusion of human intelligence with AI has the capacity to generate unique insights that can deepen our understanding of human learning processes. Through human-AI collaboration, we can design systems that are tailored to support the cognitive and metacognitive needs of individuals engaged in complex tasks. AI tools can facilitate dynamic learning environments, where humans are empowered to reflect on and regulate their thinking processes, thus enhancing their capacity for self-directed learning and collaborative problem-solving.

The synergy between human expertise and AI's ability to process large amounts of data, identify patterns, and simulate possible solutions offers the potential for novel approaches to addressing longstanding challenges in education, organizational behavior, and decision-making. Together, humans and AI can complement each other's strengths, creating dynamic feedback loops that promote continuous learning and refinement of strategies.

As AI research continues to evolve, interdisciplinary collaborations will be increasingly essential for unlocking the full potential of human-AI partnerships. Researchers who bring diverse expertise to the table will be able to explore novel methodological approaches and develop cutting-edge AI solutions that support complex collaborative activities. Furthermore, interdisciplinary collaborations can enable the identification of new research questions, methodological innovations, and sophisticated AI techniques that will drive the future of human-AI collaboration. These collaborations will be essential in designing AI systems that not only automate tasks but also enhance human capabilities—creating tools that are adaptable, scalable, and aligned with users' cognitive and socioemotional needs.

REFERENCES

1. Akata, Zeynep, Mateja Jamnik, and Allan Ramsay. "Hybrid Intelligence: Combining the Best of AI and Human Creativity." Nature Machine Intelligence 2, no. 5 (2020): 278–85. https://doi.org/10.1038/s42256-020-0198.
2. Brynjolfsson, Erik, and Andrew McAfee. The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. New York: W.W. Norton & Company, 2014.
3. Carbone, Anna, et al. "The Role of Metacognition in Human-AI Interaction: A Framework for Enhancing Collaboration." Frontiers in Artificial Intelligence 5 (2022): 120. https://doi.org/10.3389/frai.2022.00120.
4. Edwards, John D., Sarah K. Lee, and Matthew H. Miller. "Human-AI Collaboration: Designing Artificial Agents to Facilitate Socially Shared Regulation Among Learners." Journal of Learning Analytics, vol. 10, no. 2, 2023, pp. 45–67. https://doi.org/10.xxxx/jla.2023.00045.
5. Floridi, Luciano. The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities. Oxford: Oxford University Press, 2020.
6. Fong, Terry, Charles Thorpe, and Charles Baur. "Collaborative Control: A Robot-Centric Model for Vehicle Teleoperation." Industrial Robot 29, no. 4 (2002): 210–17.

7. Grosz, Barbara J. "Collaborative Systems: Leveraging Human-AI Synergy." AI Magazine 41, no. 1 (2020): 5–14. https://doi.org/10.1609/aimag.v41i1.5254.
8. Kahneman, Daniel. Thinking, Fast and Slow. New York: Farrar, Straus and Giroux, 2011.
9. Norman, Donald A. The Design of Everyday Things. Revised edition. New York: Basic Books, 2013.
10. Russell, Stuart, and Peter Norvig. Artificial Intelligence: A Modern Approach. 4th ed. Hoboken, NJ: Pearson, 2021.
11. Sarker, Md. Nazrul Islam. "Human-AI Collaboration: Exploring the Ethics of Autonomous Agents." Ethics and Information Technology 23, no. 3 (2021): 271–82. https://doi.org/10.1007/s10676-021-09581.
12. Tuunanen, Tuure, et al. "Evaluating the Effectiveness of Design Science Research in Human-AI Collaboration." Journal of Information Systems 38, no. 1 (2024): 45–61. https://doi.org/10.2139/jis2024dsr.
13. Winograd, Terry, and Fernando Flores. Understanding Computers and Cognition: A New Foundation for Design. Boston: Addison-Wesley, 1986.
14. Xu, Fei, and Frank C. Keil. "Adaptive Thinking and Human-Machine Interaction in Collaborative Environments." Cognitive Science 42, no. 8 (2021): 293–312. https://doi.org/10.1002/cogsci.2529.
15. Zuboff, Shoshana. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York: PublicAffairs, 2019.
16. Akers, John P. "Wizard of Oz Experimentation: A Tool for Understanding Human-AI Interaction." Human Factors 34, no. 2 (2023): 159–73. https://doi.org/10.1037/humaf2023oz.
17. Bentley, Peter J., and David W. Corne, eds. Creative Evolutionary Systems. Burlington, MA: Morgan Kaufmann, 2002.
18. Coeckelbergh, Mark. AI Ethics. Cambridge, MA: MIT Press, 2020.
19. Dourish, Paul. Where the Action Is: The Foundations of Embodied Interaction. Cambridge, MA: MIT Press, 2001.
20. Hollnagel, Erik. "Resilience Engineering and Human-AI Collaboration." Safety Science 120 (2020): 1–8. https://doi.org/10.1016/j.ssci.2020.104810.
21. Reeves, Byron, and Clifford Nass. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Chicago: University of Chicago Press, 1996.