INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2025, VOL. 71, NO. 4, PP. 1–9
Manuscript received April 15, 2025; revised October 2025. doi: 10.24425/ijet.2025.155474

Human-AI collaboration in Hybrid Multi-Agent Systems

Rafal Labedzki

***Abstract*—This paper examines Hybrid Multi-Agent Systems, integrating both human and non-human intelligent agents, as a new subject of management research. It presents original definitions of key concepts: intelligent agents, artificial intelligent agents, and Hybrid Multi-Agent Systems. These definitions are grounded in Distributed Artificial Intelligence and provide a foundation for exploring the collaboration between human and artificial intelligent agents. The study addresses fundamental research questions regarding the nature of intelligent agents and their role within Multi-Agent Systems, proposing Hybrid Multi-Agent Systems as a novel framework that allows for seamless cooperation between human and non-human entities. Through a narrative literature review, this paper highlights the potential implications of Hybrid Multi-Agent Systems for scientific research in management, offering a conceptual basis for future research in this evolving field.**

***Keywords*—multi-agent systems; artificial intelligence; human-AI collaboration**

I. Introduction

SINCE the 1970s, the field of Distributed Artificial Intelligence (DAI) has emerged as a significant branch of artificial intelligence, driven by the need for more capable intelligent agents. This evolution gave rise to a specialized subfield known as Multi-Agent Systems (MAS) [1].

The advent of large language models (LLMs) has revolutionized natural language processing (NLP), dramatically enhancing the communication capabilities of artificial intelligent agents. These agents can now interact not only with each other but also with humans using natural language [2], [3]. This breakthrough has redefined the conceptual framework of MAS, paving the way for the development of Hybrid Multi-Agent Systems (HyMAS). A HyMAS integrates both human and non-human intelligent agents, marking a transformative paradigm shift in the field of management.

The growing body of research underscores the increasing importance of this topic. For instance, studies such as [4]–[6] explore the collaboration between artificial intelligent agents and humans through natural language interfaces. Additionally, the relentless pursuit of artificial general intelligence (AGI) and artificial superintelligence (ASI) suggests that the so-called singularity may be approaching [7]. These advancements indicate that Hybrid Multi-Agent Systems are likely to become a reality in the near future.

R. Labedzki is with SGH Warsaw School of Economics, Warsaw, Poland (e-mail: rlabed@sgh.waw.pl).

While existing research has explored human-AI collaboration, a comprehensive theoretical framework that positions both human and artificial intelligent agents as equal participants in management processes remains absent. This paper aims to address this gap by introducing the concept of Hybrid Multi-Agent Systems. Its primary objective is to define HyMAS and examine its potential as a subject of scientific inquiry in the field of management. To this end, I propose original definitions for intelligent agents, artificial intelligent agents, Multi-Agent Systems, and Hybrid Multi-Agent Systems, using these as a foundation to discuss the implications of HyMAS for management research.

II. Method

To address the diverse issues surrounding HyMAS, including intelligent agents, MAS, NLP, and management, I adopted a narrative literature review approach. While this method lacks the methodological rigor of a systematic literature review, it provides a broad and comprehensive overview of the research topic [8]. As part of the traditional review family, the narrative literature review, despite its limited transparency, enables the synthesis of insights from extensive literature research [9].

To ensure the quality of the review, I formulated research questions [8] and designed specific literature search queries [10]. The query (”ai agent*” AND team*) was employed to search the databases EBSCO, Scopus, Emerald, and Web of Science, with the condition ”journal articles only.” This search, yielded 27, 29, 57, and 65 papers, respectively. Additionally, I utilized reference list checking [11] to further expand the scope of the literature review.

TABLE I
Research questions

ID	Research question
RQ1	What is an intelligent agent?
RQ2	What is a multi-agent system?
RQ3	Can humans be a part of a multi-agent system?
RQ4	What are the implications of multi-agent systems composed of both human and non-human agents for the scientific research in the field of management?

III. Results

A. Intelligent Agent

Although the term ”intelligent agent” has been widely used in scientific literature for many years, it’s quite challenging to

find its universally accepted definition [12]. As Carl Hewitt commented ironically at the 1994 DAI Workshop, "asking the question of what an agent is to a DAI researcher is as embarrassing as the question of what intelligence means is for an AI researcher" [13].

The origins of the concept of intelligent agent can be found in the control theory (developed by Wiener [14]), later transferred to symbolic AI and distributed AI [13]. From the perspective of the control theory, an intelligent agent would be a regulatory mechanism trying to achieve a goal state by undertaking actions that change its current state. A good example is a thermostat that measures the temperature in the room and adjusts the valve to achieve the desired temperature [12]. To be able to do it, an intelligent agent must have a perception of the environment to assess its state, a definition of a desired state, and means to change the state. Undertaking actions that lead to achieving the goals of an intelligent agent, based upon its knowledge about the environment can be understood as rational behavior [15], [16].

An intelligent agent must want to achieve its goals. Without it an agent would only know that the current state of an environment is different than desired but would not undertake any actions. This will to act is defined as an intention of an intelligent agent to achieve a desired state by using available means. That's why some researchers define intelligent agents as intentional systems [17], [18].

However, it's not enough for an intelligent agent to have means and intention to use them - agent must be allowed to do it - what leads to the need of autonomy [19]. According to Russell and Norvig [16] an autonomous intelligent agent is not only allowed to act, but also learns from its actions and improves itself.

It's also important to distinguish between means understood as actuators that an intelligent agent can use to manipulate in the environment and means understood as the resources required to use these actuators or, in the more complex systems, maintain the functioning of an intelligent agent. Resource boundedness is a key concept behind the rationality of an intelligent agent [15].

Intelligent agents function in a social space, meaning that they interact with other agents. They have to be able to take these interactions into account when planning actions that lead to achieving their goals. It's called a social ability of an agent [12].

So, what is an intelligent agent (RQ1)? It's an entity, either artificial - non-human [13], [16], [19], [20], or natural - human [15], [17], [21]–[23] that 1) is autonomous, 2) perceives its environment, 3) undertakes rational actions that are the result of this perception, its internal desires, available means and resources, as well as social constraints, and 4) learns from these actions to improve itself.

Intelligent agents can be classified according to the type of environment they inhabit. Physical agents exist in real world. They can be further divided into biological agents (for example human), artificial agents (non-human), hybrid agents (composed of biological and artificial parts). Mental agents are deployed as software systems and exist only in the virtual world [1].

B. Artificial Intelligent Agent

The term intelligent agent has been recently suppressed by the term artificial intelligence (AI) agent. In the contemporary literature it's rather hard to find notion about intelligent agents or even artificial intelligent agents, what would be correct according to the classification by Goonatilleke et al. [1]. However, there are no additional or distinctive features of artificial intelligence (AI) agents that would distinguish them from artificial intelligent agents, what allows to conclude that both terms are synonyms. In the most recent papers on AI agents it's also common to refer to them as just agents [24]. This however leads to classification on agents and humans [25]. It seems more accurate to refer to both humans and non-humans as agents [21], [23], both classes understood as intelligent agents. As a result, intelligent agents that are artificial, can be called either artificial intelligent agents, AI agents or non-human agents, whereas intelligent agents that are biological can be called biological intelligent agents, human intelligent agents or human agents. By referring to humans as human agents it's possible to naturally include them in multi-agent systems and specifically hybrid multi-agent systems that are composed of both human and non-human agents.

Artificial intelligent agents are built upon distinct architectures, which significantly influence their behavior and capabilities. These architectural differences are substantial enough to warrant the inclusion of architectural constraints in the definition of AI agents. I propose a specialized sub-definition of intelligent agents that explicitly incorporates these constraints. By introducing this definition, I aim to emphasize that the architecture of an AI agent is a critical factor that must be identified and considered when the artificial intelligent agent is the subject of scientific research.

As shown in Table II, multiple approaches to classifying AI agent architectures are well-documented in the literature. However, an analysis of the characteristics across these classes reveals that a sufficiently abstract level can be identified—one that is broad enough to encompass all classifications yet detailed enough to capture the most significant differences between them. The following analysis provides a detailed description of the selected dimensions and the approach used to determine this level of abstraction. The outcome is a high-level classification that refines the foundational definition of an intelligent agent, ultimately enabling the formulation of a specialized sub-definition for AI agent.

TABLE II
Classification of types of architectures of artificial intelligent agents

Dimension	Classes
Structure (program) type	Simple reflex agent, Model-based reflex agent, Model-based goal-based agent, Model-based utility-based agent, Hybrid (mixed)
Ability to learn	Learning, Not learning
Abstract architecture	Purely reactive, With state
Concrete architecture	Logic-based (deliberate) architectures, Reactive architectures, Belief-Desire-Intention Architectures (BDI), Layered Architectures
Behavior	Pro-active, reactive
Architecture	Reactive, Deliberative, Interacting, Hybrid

Note. Sources: [12], [13], [16]

1) Structure (program) type: All artificial intelligent agents operate in environments, that can be either physical or software-based. They perceive their environments using sensors. Actuators, like a robotic arm or a function in the code, are used by AI agents to interact with their environments. Simple reflex agents are the most basic form of artificial intelligent agents. They operate by selecting actions solely based on the current percept, without considering any history of previous perceptions. These agents rely on condition-action rules, which are essentially if-then statements that map a situation to an action. Model-based reflex agents improve upon simple reflex agents by maintaining an internal model of the world. This model allows the agent to keep track of previous percepts. By maintaining an internal state, these agents can handle partially observable environments and make decisions based on both the current percept and the internal state. Goal-based agents operate by considering future states that result from their actions and selecting actions that help them achieve their goals. Plans they formulate can contain single actions or sequences of actions. Unlike reflex agents, goal-based agents need to evaluate the desirability of different states and choose plans of actions that maximize the likelihood of reaching their goals [16]. Utility-based agents enhance goal-based agents by incorporating a utility function [26], which allows them to evaluate the desirability of different states in terms of a measure of utility. This enables the agents to not only achieve goals but also to optimize their actions according to a utility measure.

2) Ability to learn: Learning intelligent agents are designed to improve their performance over time by learning from their experiences. Learning agents often employ machine learning to improve their behavior iteratively. This architecture enables the agent to adapt to changing environments and improve its effectiveness over time [16].

3) Abstract architecture: Purely reactive agents respond to the current percept without maintaining any internal state. Their actions are based solely on the present input, making them simple but limited in dealing with dynamic or partially observable environments. Purely reactive agents implement a set of condition-action rules that map percepts directly to actions. This abstract architecture can be compared to Russell and Norvig's [16] simple reflex structure. Agents with state maintain an internal representation of the environment, allowing them to keep track of past events and use this information to inform their current decisions. State-based agents update their internal state based on new percepts and an internal model of the environment's dynamics. This state is then used to select actions. The internal state acts as a memory, providing context and continuity to the agent's actions [12]. This abstract architecture can be compared to [16] group of model-based structures.

4) Concrete architecture: Agents built using reactive architectures respond directly to percepts with minimal processing or reasoning. They can be compared to [16] simple reflex structure. Agents built using logic-based (deliberate) architectures use logical reasoning to deliberate and choose actions. They maintain an internal representation of the world and use logical inference to decide the best course of action [12]. Belief-Desire-Intention (BDI) agents are a specific type of goal-based agent that use beliefs, desires, and intentions to guide their actions [27]. Beliefs represent the information the agent has about the world, desires represent the objectives or goals the agent wants to achieve, and intentions represent the plans and actions the agent commits to in order to achieve its desires. Both BDI and deliberate architectures can be compared to Russell and Norvig's [16] goal-based and utility-based program types. Layered architectures combine multiple types of agent architectures to handle complex tasks more effectively. These architectures are typically divided into different layers, each responsible for different aspects of behavior. Commonly, there are reactive layers for immediate response, and deliberative layers for planning and reasoning. Layered architectures address resource boundedness [15], such as computational capacity, in a manner akin to the human brain's dual systems of thinking—fast and slow [28]. This approach allows the system to quickly respond to immediate, critical stimuli (fast thinking) while also enabling more complex, deliberate reasoning processes (slow thinking) when needed. This dual-layer strategy ensures both survival-critical responses and the capability for sophisticated decision-making [12].

5) Behavior: Pro-active agents take the initiative to achieve goals by planning and executing actions that lead towards desired outcomes. They are capable of anticipating future states and adjusting their behavior accordingly. Pro-active agents use planning algorithms to generate sequences of actions aimed at achieving specific goals. They typically involve goal formulation, plan generation, and plan execution phases. These agents evaluate the effects of actions on their goals and adjust their strategies to optimize outcomes [12]. Pro-active agents can be compared to Russell and Norvig's [16] goal-based and utility-based structure types. Reactive agents respond to events in the environment as they occur, without engaging in extensive planning or anticipation. Event-driven agents implement a set of event-action rules, where specific events trigger predefined actions [12]. Reactive agents can be compared either to simple reflex or model-based reflex structure described by Russell and Norvig [16].

6) Architecture: Deliberative agents rely on detailed planning and reasoning processes to make decisions. They build comprehensive models of the world, plan their actions based on these models, and execute the plans to achieve specific goals. They resemble Wooldridge's [12] logic-based (deliberative) architecture and as such can be assigned to Russell and Norvig's [16] goal-based and utility-based program types [13]. Reactive agents respond directly to environmental stimuli with pre-programmed responses. They do not engage in complex planning or reasoning but rely on condition-action rules to react to changes in their environment. They are analogous to Wooldridge's [12] reactive architectures and can be compared to Russell and Norvig's [16] simple reflex and model-based reflex structures. Hybrid agents combine elements of both deliberative and reactive architectures. They are designed to leverage the advantages of detailed planning and immediate reactivity, enabling them to operate effectively in a wide range of environments. Hybrid agents typically include a deliberative

component that handles planning and a reactive component that ensures quick responses to immediate stimuli. They can be compared to Wooldridge's [12] layered architectures. Interacting agents are designed to work together with other agents to achieve common goals. They are capable of communication, coordination, and cooperation, making them suitable for tasks that require teamwork and information sharing. Interacting agents use protocols and algorithms for negotiation, task allocation, and joint decision-making.

7) Definition of artificial intelligent agent: Dimensions analyzed above, can be reduced to one, high-level dimension. Although Russell and Norvig's [16] structure type dimension covers most of other dimensions and can be considered as a very good candidate to be the universal classification, it does not capture the resource boundedness and social interactions importance. Therefore, a reasonable approach is to use Muller's [13] architecture classification, as it generalizes all other dimensions at an appropriate level of abstraction and captures the resource boundedness and social interactions importance.

Incorporating this into the definition of an intelligent agent to create a sub-definition for an artificial intelligent agent results in the following formulation: Artificial intelligent agent is a non-human entity that is autonomous, perceives its environment, undertakes rational reactive or deliberative actions that are the result of this perception, its internal desires, available means, and resources, as well as social constraints, and learns from these actions to improve itself, while interacting with other intelligent agents.

This sub-definition emphasizes the importance of the complexity of an AI agent, that can be either reactive or deliberative or combine both. It also relates to the interacting agent architecture, highlighting the importance of the ability of AI agents to collaborate with other intelligent agents. An example of an architecture that combines reactive and deliberative components is InteRRaP presented by Muller [13]. This architecture consists of three layers: the reactive layer, the deliberative layer, and the social layer. The reactive layer handles immediate responses to environmental stimuli, while the deliberative layer focuses on planning and reasoning. The social layer facilitates communication and coordination with other agents. Figure 1a illustrates the InteRRaP architecture.

C. Multi-Agent Systems

While an artificial intelligent agent can independently take actions based on its autonomy, the true advantage of these agents is realized when they collaborate with other artificial intelligent agents. When multiple artificial intelligent agents work together to solve complex tasks, they form what is known as Multi-Agent Systems (MAS) [20]. (RQ2) MAS can be defined as a network of individual AI agents that share knowledge and communicate with each other in order to solve a problem that is beyond the scope of a single agent [29]. This definition is very close to the definition of a team (composed of human agents), that is a specialized form of task group, characterized by a small number of members whose skills complement each other to achieve a common goal that can't be achieved by any member individually [30].

To be able to collaborate, interacting AI agents [13] communicate using Agent Communication Language (ACL) [31]. Foundation of Intelligent Physical Agents (FIPA) proposed a comprehensive ACL framework for agents which has been widely used in most instantiations of MAS [20]. This language is programmatic in nature, designed to be used by non-human agents [29]. ACL aids AI agents in effectively addressing MAS challenges listed in Table III.

Although the development of ACL has been significant [32], [33], to the extent MAS can even take the form of virtual organizations and societies of agents [34], the reliance on programmatic communication has historically made these systems exclusive to non-human agents.

However, recent advancements in machine learning, particularly in NLP [35], hold the potential to revolutionize ACL by replacing programmatic communication with natural language interfaces. This shift could enable human agents to interact with non-human agents through natural language, thereby enhancing the accessibility and flexibility of MAS (RQ3). Research on integrating natural language as a common communication interface within MAS is still in its early stages [6], [36], but the goal is clear - to create systems where human and non-human agents can spontaneously form teams and coordinate shared tasks through natural language conversations.

Despite the novelty of this research area, there have already been significant developments, demonstrating that AI agents can effectively communicate using natural language [2], [3]. Moreover, these advancements extend beyond text-based communication, incorporating additional modalities such as speech [37].

This breakthrough has enabled human agents to collaborate seamlessly with non-human agents using natural language, rather than relying on traditional programmatic interfaces. Such systems are frequently referred to as hybrid human-machine systems [38], heterogeneous systems [39], or more broadly as human-AI collaboration [5], [40]. In a different context, Halloy et al. [41] introduced the term "mixed natural-artificial system" in their study on the social integration of robots within groups of cockroaches.

Building upon the presented approach that both humans and non-humans can be categorized as intelligent agents, it becomes evident that a new term is necessary to encapsulate the essence of systems that integrate both. To address this conceptual need, I propose the term "Hybrid Multi-Agent Systems" (HyMAS) to describe Multi-Agent Systems composed of both human and non-human agents.

The term "Hybrid Multi-Agent System" has previously been introduced in the context of systems composed of continuous-time and discrete-time dynamic agents, as proposed by Zheng et al. [42]. Their research primarily focused on the coordination of systems that are heterogeneous in nature, specifically within the domain of artificial intelligent agents with diverse operational features. However, their conceptualization of HyMAS does not address the collaborative interaction between human and non-human agents, but rather focuses exclusively on Multi-Agent Systems (MAS) composed solely of artificial agents with varying characteristics.

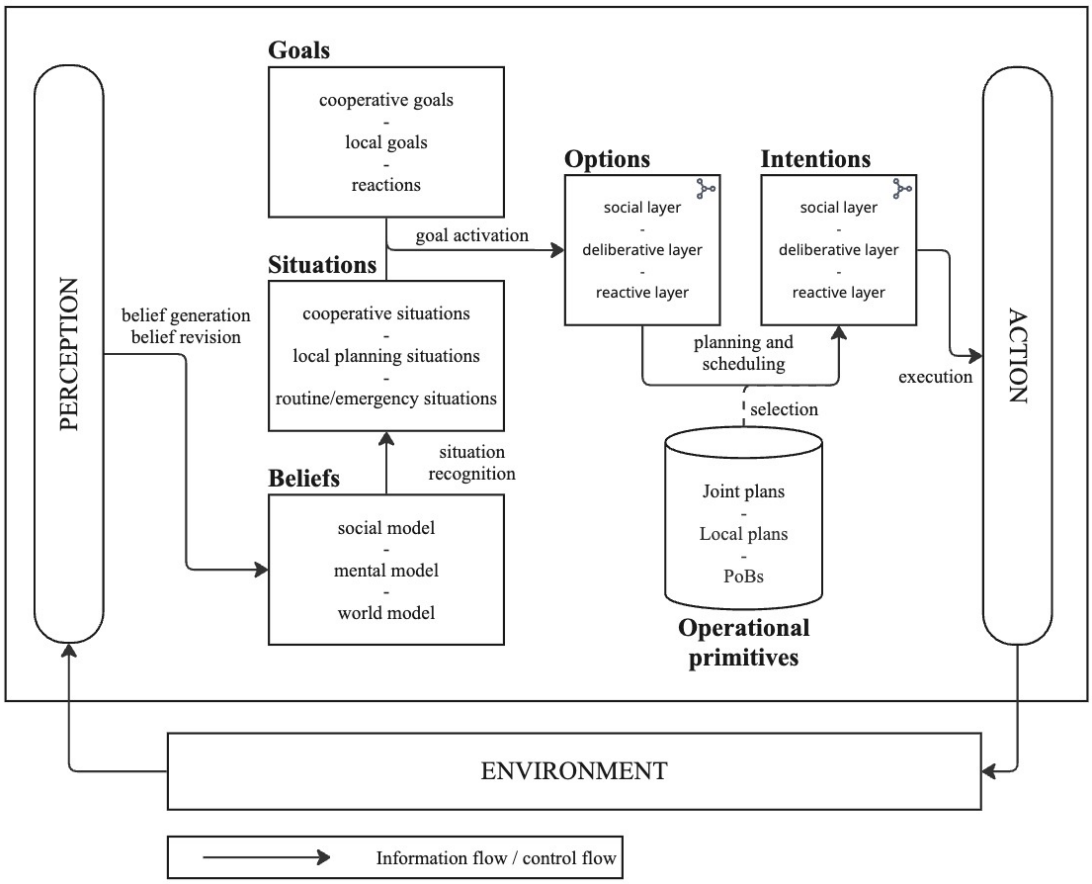

Fig. 1. The agent architecture InteRRaP

(a) InteRRaP architecture [13] is a good example of an architecture that combines reactive and deliberative components. It consists of three layers: the reactive layer, the deliberative layer, and the social layer. The reactive layer handles immediate responses to environmental stimuli, while the deliberative layer focuses on planning and reasoning. The social layer facilitates communication and coordination with other agents.

D. Management of Hybrid Multi-Agent Systems

Although the term ”Hybrid Multi-Agent System” is not entirely novel, it presents an opportunity to bridge two fields that have developed separately for over 40 years - the fields of Multi-Agent Systems and Management [1]. Management is a multidisciplinary field that involves planning, organizing, leading, and controlling an organization’s resources, including human, financial, intellectual, and physical assets, to achieve specific goals efficiently and effectively [43].

Contemporary management theory increasingly incorporates concepts of sustainability, ethics, and social responsibility. Human capital has become one of the most important competitive advantages for companies. As a result, sustainable human capital management has become increasingly important in modern organizations, emphasizing the integration of environmental and social considerations into human resources management practices to enhance job satisfaction and organizational identification across diverse cultural contexts [44].

The definition of management naturally raises the question of how the artificial components of a Hybrid Multi-Agent System (HyMAS) should be categorized - whether as subjects of human resource management, intellectual asset management, or physical asset management. Given the current stage of development of artificial intelligent agents, it would be reasonable to place them within the domains of intellectual assets or physical assets management. This categorization is justifiable by comparing existing artificial intelligent agents to software tools or machines. However, as artificial intelligence continues to advance, an increasingly pertinent question emerges: Will future AI agents still be classified merely as physical or intellectual assets, or will they transition into a category more akin to human resources within an organization?

Artificial intelligent agents may soon be viewed more as human-like resources rather than as mere components of an organization’s tangible or intangible assets [7]. As their role within HyMAS evolves from being tools to becoming team members, it is justified to assume that the primary management function in this initial stage of evolution will be organizing.

1) Organizing function: Interestingly, the organizing function has the potential to address several key challenges inherent in MAS, as outlined in Table 3. The challenges of organization, coordination, control, task allocation, and learning align closely with management processes within the organizing function [43]. The types of organizational structures employed in MAS, such as flat, hierarchical, holonic, coalition, team, and matrix organizations, are also utilized in contemporary man-

TABLE III
CHALLENGES OF MAS

Challenge	Brief description
Organization	The challenge of structuring and managing the interactions between AI agents to ensure effective collaboration and goal achievement. It refers to the way that agent communications and connections are defined. Most common organizational structures are: 1) flat, 2) hierarchical, 3) holonic, 4) coalition, 5) team, 6) matrix, 7) congregation.
Coordination control	Managing and synchronizing actions among multiple AI agents to achieve coordinated behavior and prevent conflicts or redundancy. This challenge further divides in: 1) consensus, 2) controllability, 3) synchronization, 4) connectivity, 5) formation.
Task allocation	Distributing tasks among AI agents in an efficient manner, ensuring that each agent's capabilities are utilized optimally, and tasks are completed effectively.
Learning	Enabling AI agents to adapt and improve their performance over time through experiences and interactions with the environment and other agents. Intelligent agents working together can form Multi Agent Learning (MAL) systems.
Security	Protecting the system and its communications from malicious attacks, unauthorized access, and ensuring data integrity and confidentiality.
Localization	Determining the positions of AI agents in the environment accurately, which is crucial for tasks requiring spatial awareness and navigation.
Fault detection	Identifying and addressing faults or failures in the system to maintain reliability and robustness, ensuring that the system continues to operate correctly.

Note. Source: [20]

agement practices [45]. Furthermore, the specific challenges of coordination and control, including consensus, controllability, synchronization, connectivity, and formation, mirror those encountered in management of organizations composed of people. The challenge of task allocation is central to both the organizing and planning functions of management [43]. Fields such as project management and process management address this challenge concerning human agent resources [45]. However, one of the most intriguing and potentially controversial challenges in the context of MAS and management is that of learning. Human capital management, particularly human capital development, encompasses learning as a crucial process [30]. When applied to people, learning is often a slow and resource-intensive process. In contrast, the learning process of AI agents is markedly more efficient, highlighting a potential misalignment between human and non-human actors in HyMAS. Although scientific research has thus far focused on issues such as trust between humans and AI [4], [46]–[48] or the perception of AI as a virtual assistant [49], [50], it is likely that the evolution of intelligent agents within HyMAS, particularly those built with deliberative architecture, will bring new alignment challenges, like that of learning.

2) Leading function: As AI will continue to evolve, the growing importance of the leading function within management will become apparent. Although ethically controversial, there is a non-zero probability that, at some point, non-human members of HyMAS will attain a level of autonomy that necessitates leadership directed toward them. The concept of leadership within HyMAS could become even more controversial as it raises questions about who the leader is and who follows [1]. This leads to the critical inquiry of whether an AI agent can be considered a leader of human agents within a HyMAS. If the answer is affirmative, the leading function of management would require a profound rethinking. Even in scenarios where AI agents remain followers rather than leaders, human leaders will need to adapt to this new reality, potentially leading to modifications in the traditional leading function. For instance, the managerial grid proposed by Blake and Mouton [51] may require the addition of a third dimension - AI agent orientation.

3) Controlling function: Similar considerations arise when examining the controlling function. While it is relatively easy to envision maintaining control over reactive agents, that function more as tools, the challenge becomes more complex when dealing with deliberative, interacting agents granted high levels of autonomy, effectively making them equal to human members of the HyMAS. Traditionally, the controlling function is grounded in values that form the foundation of an organization's culture. However, the emergence of HyMAS introduces actors into the organization that, at least for now, cannot be easily associated with any particular set of values. Future research in management will need to investigate not only how to effectively control artificial intelligent agents within HyMAS, but also how these agents can exert control over both other AI agents and human agents.

4) Planning function: Machine learning currently supports the planning and decision-making function within organizations, enhancing efficiency by processing large datasets and generating insights that inform choices. However, this increased reliance on AI has sparked controversies, particularly regarding algorithmic discrimination, where biases embedded in the data can lead to unfair outcomes [52]. Another significant concern is the explainability of AI systems, as decision-makers and stakeholders often demand transparency in understanding how AI models arrive at their conclusions [53]–[56]. The introduction of HyMAS, where artificial intelligent agents are empowered to participate in the planning function, further complicates this dynamic. AI agents' involvement in the planning function of management alongside humans, can raise questions about accountability, bias amplification, and the ethical implications.

5) MAS specific challenges: The introduction of non-human agents into organizations introduces challenges that are distinct from those associated with human agents. As outlined in Table III, challenges such as security, localization, and fault detection are specific to artificial intelligent agents. If these agents are integrated into an organization as part of a HyMAS, management must also address these challenges.

IV. Implications for Management Research

The concept of an organization as a HyMAS represents a novel and complex subject for the field of management. The introduction of artificial intelligent agents, previously unencountered actors in organizational settings, fundamentally alters the traditional management paradigm, which has long been based on the management of people by people. Organizations that have traditionally concentrated on the attributes

of human agents will have to consider non-human factors as integral components of their operations. This challenge must be addressed by scientific research in the field of management.

To explore the implications of Multi-Agent Systems composed of both human and non-human agents for scientific research in the field of management (RQ4), several key issues must be considered:

- Novelty and technical nature of the subject: The study of HyMAS, particularly when it includes artificial intelligent agents, is inherently novel and technical. As highlighted in this paper, AI agents can be constructed using various architectures, which directly influence the type of entity they become. When such agents are the focus of management research, it is imperative to address these technical constraints to ensure the replicability of research and the reliability of its results.
- Current nonexistence of fully developed HyMAS: As of the time of writing, fully developed HyMAS, where human and non-human agents operate as equal members, do not yet exist. The future existence of such systems remains uncertain. However, management theory must be prepared to address the challenges these systems will present if and when they become reality. To meet this requirement, scientific research must employ methods that can keep pace with the rapid development of the subject. It is highly probable that during scientific research with HyMAS, the rapid evolution of AI agents will render some research assumptions obsolete, necessitating a dynamic and adaptable approach.
- Misalignment between management and AI research: There is a visible misalignment between research in management and in artificial intelligence. Contemporary management research often equates AI with machine learning, leading to a proliferation of studies focused on the use of machine learning tools within organizations. This perspective tends to support the conclusion that AI can only augment human work, remaining complementary rather than substitutive to the human workforce. Simultaneously, there is a significant body of research in the field of distributed artificial intelligence, with numerous studies and grey literature published each month. These works often present a different perspective, with some aiming for artificial general intelligence (AGI) and eventually artificial superintelligence [57].
- The changing relevance of the Moravec Paradox: The Moravec Paradox, which posits that high-level reasoning requires less computational power than low-level sensorimotor skills [58], is becoming less applicable. The ability to train AI agents in virtual environments, allowing them to undergo the equivalent of thousands of years of training, has resulted in skills that surpass those of any human [59]. With national governments now engaged in an accelerated race toward AGI, it is highly likely that HyMAS will become a reality in the near future.

Hybrid Multi-Agent Systems have the potential to become a reality in the near future. However, it is important to acknowledge that, at present, they remain largely theoretical. The integration of Large Language Models into AI agent architectures is still in its nascent stages. Even the most advanced LLM-based tools currently function most often as reactive agents, primarily responding to user inputs. Nonetheless, this situation is expected to evolve rapidly. Leading companies in the AI sector, including OpenAI, X.com, Anthropic, Meta and Google, have made it clear through their predictions and ongoing research that significant advancements are imminent. Their focus on topics such as Universal Basic Income (UBI) underscores the anticipation of a post-AGI economy, where widespread automation may lead to significant unemployment, necessitating innovative solutions like UBI to address these challenges [60].

V. Conclusions

There is a prevalent narrative within the scientific community that strongly advocates for the notion that the only viable path for human-AI collaboration involves augmenting human capabilities with machine intelligence. This perspective, while influential, represents just one of several potential models for future collaboration. With this study, I propose an alternative approach in which humans are conceptualized as human agents and integrated into Hybrid Multi-Agent Systems on equal footing with non-human agents. This framework allows for the exploration of a broader spectrum of collaboration types, providing a more comprehensive understanding of potential human-machine interactions. Moreover, it offers a basis for developing management knowledge that addresses the diverse possibilities inherent in these interactions.

As AI continues to advance rapidly, management research must evolve to remain relevant and capable of addressing the complexities of future organizations where human and non-human agents collaborate. The first critical step is revising and refining the methodologies used in management research to accommodate this emerging reality. By employing up-to-date scientific methods, researchers will be better equipped to address the unique challenges posed by Hybrid Multi-Agent Systems and their implications for management theory and practice.

References

[1] S. Goonatilleke and B. Hettige, "Past, Present and Future Trends in Multi-Agent System Technology," *Journal Européen des Systèmes Automatisés*, vol. 55, pp. 723–739, Dec. 2022.

[2] S. Han, Q. Zhang, Y. Yao, W. Jin, Z. Xu, and C. He, "LLM Multi-Agent Systems: Challenges and Open Problems," Feb. 2024, arXiv:2402.03578 [cs]. [Online]. Available: http://arxiv.org/abs/2402.03578

[3] Y. Talebirad and A. Nadiri, "Multi-Agent Collaboration: Harnessing the Power of Intelligent LLM Agents," Jun. 2023, arXiv:2306.03314 [cs]. [Online]. Available: http://arxiv.org/abs/2306.03314

[4] A. R. Dennis, A. Lakhiwal, and A. Sachdeva, "AI Agents as Team Members: Effects on Satisfaction, Conflict, Trustworthiness, and Willingness to Work With," *Journal of Management Information Systems*, vol. 40, no. 2, pp. 307–337, Apr. 2023, publisher: Routledge. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/07421222.2023.2196773

[5] G. Zhang, L. Chong, K. Kotovsky, and J. Cagan, "Trust in an AI versus a Human teammate: The effects of teammate identity and performance on Human-AI cooperation," *Computers in Human Behavior*, vol. 139, 2023. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85140287945&doi=10.1016%2fj.chb.2022.107536&partnerID=40&md5=076843351fbd906c0c583f0895f313b5

[6] O. Lemon, "Conversational AI for multi-agent communication in Natural Language," *AI Communications*, vol. 35, no. 4, pp. 295–308, 2022. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85140849537&doi=10.3233%2fAIC-220147&partnerID=40&md5=50418fa9218a4a660ec6578da3fe4de6

[7] L. Aschenbrenner, "Situational Awareness," 2024. [Online]. Available: https://situational-awareness.ai

[8] J. A. Collins and B. C. Fauser, "Balancing the strengths of systematic and narrative reviews," *Human Reproduction Update*, vol. 11, no. 2, pp. 103–104, Mar. 2005. [Online]. Available: http://academic.oup.com/humupd/article/11/2/103/763121/Balancing-the-strengths-of-systematic-and

[9] A. Sutton, M. Clowes, L. Preston, and A. Booth, "Meeting the review family: exploring review types and associated information retrieval requirements," *Health Information & Libraries Journal*, vol. 36, no. 3, pp. 202–222, 2019, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/hir.12276. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/hir.12276

[10] J. A. Byrne, "Improving the peer review of narrative literature reviews," *Research Integrity and Peer Review*, vol. 1, no. 1, p. 12, Sep. 2016. [Online]. Available: https://doi.org/10.1186/s41073-016-0019-2

[11] T. Horsley, O. Dingwall, and M. Sampson, "Checking reference lists to find additional studies for systematic reviews," *Cochrane Database of Systematic Reviews*, no. 8, 2011, publisher: John Wiley & Sons, Ltd. [Online]. Available: https://doi.org//10.1002/14651858.MR000026.pub2

[12] M. Wooldridge, "Intelligent Agents: The Key Concepts," in *Multi-Agent Systems and Applications II*, V. Mařík, O. Štěpánková, H. Krautwurmová, and M. Luck, Eds. Berlin, Heidelberg: Springer, 2002, pp. 3–43.

[13] J. P. Müller, *The design of intelligent agents: a layered approach*, ser. Lecture notes in computer science ; Lecture notes in artificial intelligence. Berlin ; New York: Springer, 1996, no. 1177.

[14] N. Wiener, *Cybernetics or control and communication in the animal and the machine*, 2nd ed. Cambridge, Mass: MIT Press, 2000.

[15] M. E. Bratman, D. J. Israel, and M. E. Pollack, "Plans and resource-bounded practical reasoning," *Computational Intelligence*, vol. 4, no. 3, pp. 349–355, Sep. 1988. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/j.1467-8640.1988.tb00284.x

[16] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach*, third edition, global edition ed., ser. Prentice Hall series in artificial intelligence. Boston Columbus Indianapolis New York San Francisco Upper Saddle River Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montreal Toronto Delhi Mexico City Sao Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo: Pearson, 2016.

[17] D. C. Dennett, "Intentional Systems," *The Journal of Philosophy*, vol. 68, no. 4, pp. 87–106, 1971, publisher: Journal of Philosophy, Inc. [Online]. Available: https://www.jstor.org/stable/2025382

[18] N. Seel, *AGENT THEORIES AND ARCHITECTURES*. London: STC PLC: STC Technology Ltd, 1989.

[19] M. Wooldridge and N. Jennings, "Intelligent agents: theory and practice," *The Knowledge Engineering Review*, 1995. [Online]. Available: https://www.semanticscholar.org/author/M.-Wooldridge/48106342

[20] A. Dorri, S. S. Kanhere, and R. Jurdak, "Multi-Agent Systems: A Survey," *IEEE Access*, vol. 6, pp. 28 573–28 593, 2018, conference Name: IEEE Access. [Online]. Available: https://ieeexplore.ieee.org/document/8352646

[21] M. J. Casper, "Reframing and Grounding Nonhuman Agency: What Makes a Fetus an Agent," *American Behavioral Scientist*, vol. 37, no. 6, pp. 839–856, May 1994. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0002764294037006009

[22] M. Emirbayer and A. Mische, "What Is Agency?" *American Journal of Sociology*, vol. 103, no. 4, pp. 962–1023, Jan. 1998. [Online]. Available: https://www.journals.uchicago.edu/doi/10.1086/231294

[23] W. Rammert, "Distributed Agency and Advanced Technology Or: How to Analyse Constellations of Collective Inter-Agency," 2012.

[24] L. Fan, G. Wang, Y. Jiang, A. Mandlekar, Y. Yang, H. Zhu, A. Tang, D.-A. Huang, Y. Zhu, and A. Anandkumar, "MINEDOJO: Building Open-Ended Embodied Agents with Internet-Scale Knowledge," 2022.

[25] R. Dunning, B. Fischhoff, and A. Davis, "When Do Humans Heed AI Agents' Advice? When Should They?" *Human Factors*, 2023. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85167579278&doi=10.1177%2f00187208231190459&partnerID=40&md5=784bb8e028c3e47febe036fcbc57f929

[26] J. Eatwell, M. Milgate, and P. Newman, Eds., *Utility and Probability*. London: Palgrave Macmillan UK, 1990. [Online]. Available: http://link.springer.com/10.1007/978-1-349-20568-4

[27] A. S. Rao and M. P. Georgeff, "Modeling Rational Agents within a BDI-Architecture," 1991, publisher: Citeseer.

[28] D. Kahneman, *Thinking, Fast and Slow*. Penguin UK, Nov. 2011, google-Books-ID: oV1tXT3HigoC.

[29] B. Parasumanna Gokulan and D. Srinivasan, "An Introduction to Multi-Agent Systems," in *Studies in Computational Intelligence*, Jul. 2010, vol. 310, pp. 1–27, journal Abbreviation: Studies in Computational Intelligence.

[30] M. Armstrong, "Zarzadzanie zasobami ludzkimi, Oficyna Ekonomiczna, Kraków," *Search in*, p. 245, 2005.

[31] S. Poslad, "Specifying protocols for multi-agent systems interaction," *ACM Transactions on Autonomous and Adaptive Systems*, vol. 2, no. 4, p. 15, Nov. 2007. [Online]. Available: https://dl.acm.org/doi/10.1145/1293731.1293735

[32] C. Castelfranchi and Y. Lespérance, Eds., *Intelligent agents VII: agent theories architectures and languages: 7th International Workshop, ATAL 2000, Boston, MA, USA, July 7-9, 2000: proceedings*, ser. Lecture notes in computer science ; Lecture notes in artificial intelligence. Berlin ; New York: Springer, 2001, no. 1986, meeting Name: ATAL 2000.

[33] B. Chaib-draa and F. Dignum, "Trends in Agent Communication Language," *Computational Intelligence*, vol. 18, no. 2, pp. 89–101, 2002, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8640.00184. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8640.00184

[34] Jose A. Maestro-Prietoa and Sara Rodrigue, "Agent organisations: from independent agents to virtual organisations and societies of agents," *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 9, no. 4, 2020.

[35] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, u. Kaiser, and I. Polosukhin, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[36] A. Lazaridou, A. Potapenko, and O. Tieleman, "Multi-agent Communication meets Natural Language: Synergies between Functional and Structural Language Learning," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, Eds. Online: Association for Computational Linguistics, Jul. 2020, pp. 7663–7674. [Online]. Available: https://aclanthology.org/2020.acl-main.685

[37] D. Zhang, Z. Li, P. Wang, X. Zhang, Y. Zhou, and X. Qiu, "SpeechAgents: Human-Communication Simulation with Multi-Modal Multi-Agent Systems," Jan. 2024, arXiv:2401.03945 [cs]. [Online]. Available: http://arxiv.org/abs/2401.03945

[38] I. Rahwan, M. Cebrian, N. Obradovich, J. Bongard, J.-F. Bonnefon, C. Breazeal, J. W. Crandall, N. A. Christakis, I. D. Couzin, M. O. Jackson, N. R. Jennings, E. Kamar, I. M. Kloumann, H. Larochelle, D. Lazer, R. McElreath, A. Mislove, D. C. Parkes, A. entland, M. E. Roberts, A. Shariff, J. B. Tenenbaum, and M. Wellman, "Machine behaviour," *Nature*, vol. 568, no. 7753, pp. 477–486, Apr. 2019. [Online]. Available: https://www.nature.com/articles/s41586-019-1138-y

[39] H. Shirado and N. A. Christakis, "Locally Noisy Autonomous Agents Improve Global Human Coordination in Network Experiments," *Nature*, vol. 545, no. 7654, pp. 370–374, May 2017. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5912653/

[40] R. Jain, N. Garg, and S. N. Khera, "Effective human–AI work design for collaborative decision-making," *Kybernetes*, vol. 52, no. 11, pp. 5017–5040, Jan. 2023, publisher: Emerald Publishing Limited. [Online]. Available: https://doi.org/10.1108/K-04-2022-0548

[41] J. Halloy, G. Sempo, G. Caprari, C. Rivault, M. Asadpour, F. Tâche, I. Saïd, V. Durier, S. Canonge, J. M. Amé, C. Detrain, N. Correll, A. Martinoli, F. Mondada, R. Siegwart, and J. L. Deneubourg, "Social Integration of Robots into Groups of Cockroaches to Control Self-Organized Choices," *Science*, vol. 318, no. 5853, pp. 1155–1158, Nov. 2007, publisher: American Association for the Advancement of Science. [Online]. Available: https://www.science.org/doi/10.1126/science.1144259

[42] Y. Zheng, Q. Zhao, J. Ma, and L. Wang, "Second-order consensus of hybrid multi-agent systems," *Systems & Control Letters*, vol. 125, pp. 51–58, Mar. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167691119300179

[43] R. W. Griffin, *Management*, 11th ed. Australia ; Mason, OH: South-Western Cengage Learning, 2013.

[44] A. Wojtczuk-Turek, D. Turek, F. Edgar, H. J. Klein, J. Bosak, B. Okay-Somerville, N. Fu, S. Raeder, P. Jurek, and A. Lupina-Wegener, "Sustainable human resource management and job satisfaction—Unlocking

the power of organizational identification: A cross-cultural perspective from 54 countries," *Corporate Social Responsibility and Environmental Management*, 2024, publisher: Wiley Online Library.

[45] H. Kerzner, *Advanced project management: Best practices on implementation*. John Wiley & Sons, 2003.

[46] B. Gebru, L. Zeleke, D. Blankson, M. Nabil, S. Nateghi, A. Homaifar, and E. Tunstel, "A Review on Human–Machine Trust Evaluation: Human-Centric and Machine-Centric Perspectives," *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 5, pp. 952–962, Oct. 2022, conference Name: IEEE Transactions on Human-Machine Systems. [Online]. Available: https://ieeexplore.ieee.org/document/9720720

[47] Z. R. Khavas, S. R. Ahmadzadeh, and P. Robinette, "Modeling Trust in Human-Robot Interaction: A Survey," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12483 LNAI, pp. 529–541, 2020, arXiv: 2011.04796 Publisher: Springer Science and Business Media Deutschland GmbH ISBN: 9783030620554.

[48] A.-S. Ulfert, E. Georganta, C. Centeio Jorge, S. Mehrotra, and M. Tielman, "Shaping a multidisciplinary understanding of team trust in human-AI teams: a theoretical framework," *European Journal of Work and Organizational Psychology*, 2023. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85153532721&doi=10.1080%2f1359432X.2023.2200172&partnerID=40&md5=aa44582a54605f3ab7c7332904e38f7e

[49] L. Ciechanowski, A. Przegalinska, M. Magnuski, and P. Gloor, "In the shades of the uncanny valley: An experimental study of human–chatbot interaction," *Future Generation Computer Systems*, vol. 92, pp. 539–548, Mar. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167739X17312268

[50] A. Przegalinska, L. Ciechanowski, A. Stroz, P. Gloor, and G. Mazurek, "In bot we trust: A new methodology of chatbot performance measures," *Business Horizons*, vol. 62, no. 6, pp. 785–797, Nov. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0007681319301177X

[51] R. R. Blake and J. S. Mouton, *The new managerial grid : strategic new insights into a proven system for increasing organization productivity and individual effectiveness, plus a revealing examination of how your managerial style can affect your mental and physical health*. Houston : Gulf Pub. Co., Book Division, 1978. [Online]. Available: http://archive.org/details/newmanagerialgri00blak

[52] M. Wójcik, "Algorithmic discrimination in the era of artificial intelligence: challenges of sustainable human resource management," *Edukacja Ekonomistów i Menedżerów*, vol. 69, no. 3, 2023, number: 3. [Online]. Available: https://econjournals.sgh.waw.pl/EEiM/article/view/4540

[53] B. Abedin, "Managing the tension between opposing effects of explainability of artificial intelligence: a contingency theory perspective," *Internet Research*, vol. 32, no. 2, pp. 425–453, Jan. 2022, publisher: Emerald Publishing Limited. [Online]. Available: https://doi.org/10.1108/INTR-05-2020-0300

[54] D. Gunning, E. Vorm, J. Y. Wang, and M. Turek, "DARPA's explainable AI (XAI) program: A retrospective," *Applied AI Letters*, vol. 2, no. 4, p. e61, 2021, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ail2.61. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ail2.61

[55] A. Picard, Y. Mualla, F. Gechter, and S. Galland, "Human-Computer Interaction and Explainability: Intersection and Terminology," vol. 1902 CCIS, 2023, pp. 214–236. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85176005051&doi=10.1007%2f978-3-031-44067-0_12&partnerID=40&md5=dbd4fe92a6e5d28a51536ef1d6670209

[56] A. Rosenfeld and A. Richardson, "Explainability in human–agent systems," *Autonomous Agents and Multi-Agent Systems*, vol. 33, no. 6, pp. 673–705, Nov. 2019. [Online]. Available: https://doi.org/10.1007/s10458-019-09408-y

[57] T. Mulgan, "Superintelligence: Paths, dangers, strategies," 2016, publisher: Oxford University Press.

[58] J. E. H. Korteling, G. C. van de Boer-Visschedijk, R. A. M. Blankendaal, R. C. Boonekamp, and A. R. Eikelboom, "Human-versus Artificial Intelligence," *Frontiers in Artificial Intelligence*, vol. 4, Mar. 2021, publisher: Frontiers Media S.A. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frai.2021.622364/full

[59] "NVIDIA Isaac Sim · GitHub," 2024. [Online]. Available: https://github.com/isaac-sim

[60] "Sam Altman-Backed Group Completes Largest US Study on Basic Income," *Bloomberg.com*, Jul. 2024. [Online]. Available: https://www.bloomberg.com/news/articles/2024-07-22/ubi-study-backed-by-openai-s-sam-altman-bolsters-support-for-basic-income