Multi-Agent AI

Simeon Allmendinger[1,2*], Lukas Bonenberger[4], Kathrin Endres[2,3], Dominik Fetzer[2,3], Henner Gimpel[2,3], Niklas Kühl[1,2]

[1]Information Systems and Human-Centric Artificial Intelligence, University of Bayreuth, Bayreuth, Germany.
[2]Business & Information Systems Engineering, Fraunhofer-Institute for Applied Information Technology, Bayreuth, Germany.
[3]Digital Management, University of Hohenheim, Stuttgart, Germany.
[4]Trusteq GmbH, Munich, Germany.

*Corresponding author(s). E-mail(s): simeon.allmendinger@uni-bayreuth.de;

Abstract

Multi-Agent Artificial Intelligence (MAAI) represents a foundational shift in the automation of knowledge work, moving beyond static workflows toward adaptive systems of interacting AI-based agents. These agents perceive, reason, and coordinate in real time to address complex, context-rich tasks that traditionally require human expertise. Drawing on the conceptual roots of process automation, agentic information systems, and AI, this paper introduces a structured, five-component framework that conceptualizes MAAI as a layered architecture based on: foundation model, data-centric perception and action, dynamic orchestration, agent-integrated workflow, and interaction interface. This framework disentangles the technical, organizational, and human-facing dimensions of MAAI, offering researchers and practitioners a systematic lens to analyze and design agent-based AI automation. The framework further structures three research pathways focused on advancing technical capabilities, enabling organizational integration, and addressing socio-technical implications such as fairness, accountability, and labor transformation. Together, these contributions establish a foundation for interdisciplinary inquiry into how MAAI reshapes work, coordination, and digital value creation.[*]

Keywords: Agentic Information Systems, Artificial Intelligence, Process Automation, Multi-Agent-System

[*]Discussion paper available at https://osf.io/hndm3/

1 Introduction

Imagine a scenario in which a customer contacts their insurance provider to inquire whether recent damage to their vehicle is covered under their current policy. The inquiry—submitted via a customer portal or mobile app—includes a short textual description and photographs of the damage. This seemingly simple request triggers a series of internal processing steps within the insurance company. A customer service representative must interpret the inquiry and extract relevant details. A policy analyst then needs to assess whether the reported case aligns with the customer's contract terms, potentially consulting internal databases or regulatory frameworks. Subsequently, a claims processor cross-checks the case with historical claims and pricing information to evaluate plausibility and prepare a response. Each of these roles contributes expert knowledge and judgment to tailor a case-specific answer. This scenario clearly illustrates that knowledge work involves not merely processing information but creating, distributing, and applying (expert) knowledge. Skilled individuals autonomously manage information and data to navigate complex situations and produce tailored outcomes (Drucker, 1999; Pyöriä, 2005; Schultze, 2000). This type of knowledge work requires a deep understanding of framed experience, values, contextual information, and expert insight. Accordingly, the tasks involved are typically unstructured, non-routine, and intricately tailored to specific situations (Davenport, Jarvenpaa, & Beers, 1996; Davenport & Prusak, 1998; Heerwagen, Kampschroer, Powell, & Loftness, 2004; Janz, Conquitt, & Noe, 1997).

With the advancement of artificial intelligence (AI), automation in information systems (IS) is entering a new phase. While the integration of AI into business processes has already streamlined routine tasks, we now see a shift toward systems that not only execute, but also reason, interact, and adapt (Acharya, Kuppan, & Divya, 2025). The increasing availability of AI capabilities in the form of AI-as-a-Service (AIaaS) drives this evolution (Lins et al., 2021). However, greater advancements result from the emergence of a new phenomenon: *Multi-Agent AI* (MAAI). MAAI marks a fundamental departure from traditional forms of automation. Rather than following predefined workflows, MAAI is composed of multiple AI-based agents—each capable of perception, reasoning, and decision-making (W. Zhou et al., 2023)—that dynamically coordinate, delegate, and respond to evolving environments (Baird & Maruping, 2021). They are designed to tackle unstructured, context-dependent tasks through collaboration and self-organization, thereby introducing a new layer of cognitive capability into IS. In contrast to the rigid logic of robotic process automation (RPA) (Hofmann, Samp, & Urbach, 2020; Jimenez-Ramirez, Reijers, Barba, & Del Valle, 2019), MAAI exhibits flexible, emergent behavior suited for complex knowledge work. The rise of low-code tools (Bock & Frank, 2021) that lower the technical hurdles for developing and deploying agent-based systems at scale further underscores the practical relevance of MAAI. Tools and platforms such as AutoGen[1], CrewAI[2], Swarm[3], and the Bee

[1] https://microsoft.github.io/autogen/0.2/
[2] https://www.crewai.com
[3] https://github.com/openai/swarm

Agent Framework[4] illustrate how MAAI is rapidly transitioning from experimental technology to accessible infrastructure for real-world applications.

However, as MAAI expands the horizons of what AI-based IS can do, it also introduces new layers of complexity and ambiguity—rooted not only in its technical conceptualization but also in the conceptual tensions inherited from its origins in process automation, agentic IS, and AI. These challenges pertain not only to the underlying technological complexity of MAAI but also to its far-reaching implications for work structures, occupational roles, and digital value creation, as it gradually enables the substitution of human labor through more intelligent, adaptive, and context-aware forms of automation (Svanberg, Li, Fleming, Goehring, & Thompson, 2024). To bring conceptual clarity and theoretical grounding to this rapidly evolving debate, we introduce a foundational framework that structures MAAI along five interdependent components: (1) *foundation model (FM)*, (2) *data-centric perception and action*, (3) *dynamic orchestration*, (4) *agent-integrated workflow*, and (5) *interaction interface*. These components serve as layered debate spaces—Matrjoschka-like conceptual containers—that help order the discourse on MAAI within the IS field. By disentangling these layers, our MAAI framework not only explains *what MAAI is* and *how MAAI works*, but also provides the fundamentals needed to *systematically investigate MAAI further*. With our framework pinpointing central concepts per MAAI component, we provide a structured basis for the Electronic Markets (EM) community and other IS research communities to engage across emerging pathways, including technical capabilities such as FMs (Schneider, Meske, & Kuss, 2024) and generative AI (Banh & Strobel, 2023; Feuerriegel, Hartmann, Janiesch, & Zschech, 2024), organizational and economic integration such as human-AI collaboration (Guo et al., 2024; Song, Tan, Zhu, Feng, & Lee, 2024), and socio-technical implications such as appropriate reliance (Schemmer, Kuehl, Benz, Bartos, & Satzger, 2023). By clearly delineating these conceptual domains in an overarching framework, it establishes a foundation for developing targeted research avenues within each component and supports a more focused and coherent discourse on MAAI within the IS community. As novel MAAI developments (Acharya et al., 2025) accelerate, MAAI holds the potential not only to restructure knowledge work but also to redefine occupational roles and reconfigure digital markets. Against this backdrop, grounding MAAI in its conceptual roots becomes essential for making sense of its complexity and ensuring that the field of IS can engage with it through shared language, structured analysis, and theoretically informed debate.

In the following sections, we provide a structured exploration of the MAAI landscape. Section 2 examines the conceptual roots of MAAI, offering theoretical grounding that situates the phenomenon within existing strands of IS research. Section 3 introduces our framework for MAAI, articulating five interrelated components that collectively define the core layers of AI-based multi-agent systems. Section 4 then outlines key implementation challenges and identifies emerging research pathways, demonstrating how the proposed framework enables a systematic investigation of MAAI's technical capabilities, organizational and economic integration, and socio-technical implication. Together, these sections establish a foundational understanding of MAAI

[4]https://github.com/i-am-bee/bee-agent-framework

as both a conceptual innovation and a transformative force in the evolving field of knowledge work automation.

2 Conceptual Roots

The classification of systems as MAAI is debated in IS research, due to ambiguities in their conceptual boundaries and inconsistencies in theoretical grounding. In technical terms, MAAI typically refers to systems comprising multiple interacting AI agents, each with autonomy and decision-making capacity (Hughes et al., 2025), consistent with the multi-agent systems (MAS) tradition in computer science, where agents coordinate or compete to achieve goals (Ferber & Weiss, 1999; Wooldridge, 2009). However, within the IS domain, the boundaries are less clear (Dorri, Kanhere, & Jurdak, 2018). For instance, it remains debated whether complex software with multiple (generative) AI modules already qualifies as a MAS, or whether distinct agent identities and goals are required. This ambiguity has prompted a re-evaluation of core IS assumptions. To address this, MAAI can be situated at the intersection of three foundational IS research streams, which are widely applied in digital market ecosystems: process automation (1), agentic IS (2), and AI (3).

Process automation traditionally involves transferring work that is organized into processes consisting of events, tasks, and decision points from humans to technology (Engel, Ebel, & Leimeister, 2022; Ye et al., 2023), often to enhance productivity and reduce manual workload (Autor, Levy, & Murnane, 2003; Ye et al., 2023). Robotic process automation (RPA) exemplifies this, using predefined business rules to automate routine tasks (Hofmann et al., 2020; "IEEE Guide for Terms and Concepts in Intelligent Process Automation", 2017; Jimenez-Ramirez et al., 2019). It is especially effective in repetitive tasks (Haase et al., 2024; Willcocks, Lacity, & Craig, 2017). Yet, traditional IS theory often framed such automation technologies as passive artifacts under human control (Baird & Maruping, 2021). This framing holds for classical RPA systems, which do not possess autonomy or adaptive behavior. However, as automation increasingly incorporates AI capabilities, this paradigm is challenged.

The second foundational stream informing MAAI is **agentic IS** (Anonymous, 2025). Reflecting on agentic IS, Baird and Maruping (2021) developed the IS Delegation Framework. This framework defines agentic IS through three constructs: autonomous agents, their relational mechanisms (e.g., delegation), and the tasks they collaboratively pursue. At the core of agentic IS are agents, human or computational, which are autonomous entities capable of perceiving and acting within an environment, making independent decisions, and interacting with each other (Russell & Norvig, 2016). In doing so, they cooperate and occasionally conflict to achieve individual or collective goals (Baird & Maruping, 2021). Agentic IS artifacts challenge traditional IS views of passive technologies by autonomously initiating actions, managing tasks, and assuming roles and responsibilities (Baird & Maruping, 2021; Feuerriegel et al., 2024). In this context, Baird and Maruping (2021) propose rethinking IS theory to incorporate AI artifacts as agentic actors. Similarly, Schuetz and Venkatesh (2020) argue that cognitive, agentic systems exceed the explanatory scope of existing models of user-IT interaction, as they can adapt, perceive their environment, and interact with humans and technologies. While Baird and Maruping (2021) primarily focus on dyads of human and agentic IS artifacts, configurations involving multiple interacting artifacts align with MAS (Dorri et al., 2018). However, definitional ambiguity persists. Some scholars treat agentic systems broadly, encompassing single- and multi-agent setups (Baird & Maruping, 2021), while others reserve "multi-agent" for systems with multiple AI agents operating in collaboration (Hughes et al., 2025). This has led to inconsistent terminology across the literature. Baird and Maruping (2021) use the term "agentic IS artifacts", while Schuetz and Venkatesh (2020) prefer "cognitive computing systems". Other works refer to "AI agents" or "MAS"(Göldi & Rietsche, 2024; Hughes et al., 2025), or to more specific labels such as "large language model-based agents"(Xi et al., 2025), contributing to conceptual ambiguity, communication challenges, and fragmentation across IS discourse.

The third foundational stream underpinning MAAI is **AI**, which has significantly advanced both process automation and agentic IS. While RPA has already introduced considerable value in digital markets through its application in automating rule-based, repetitive tasks, its scope remains limited to non-cognitive activities. To overcome these limitations, several attempts have been made to incorporate AI into process automation (Engel et al., 2022). AI represents "techniques facilitating machines to mimic human behavior, i.e., reproducing or excelling over human decision-making with the goal of solving complex tasks" (Engel et al., 2022, p. 342). The integration of AI into process automation approaches extends the automatable business processes beyond deterministic routines (Janiesch, Zschech, & Heinrich, 2021). Traditional machine learning methods introduce probabilistic decision-making into automation, giving rise to "cognitive automation"(Engel et al., 2022; Hofmann et al., 2020). For instance, AI can be used to route emails based on content classification, with subsequent execution handled via RPA (Engel et al., 2022). Similarly, agentic IS also benefits from advancements in AI, which augment the autonomy and intelligence of agentic IS artifacts (Feuerriegel et al., 2024; Kuehl, Schemmer, Goutier, & Satzger, 2022). AI technologies equip software entities with cognitive functions, enabling them to perform complex knowledge work traditionally requiring human judgment (Berente, Gu, Recker, & Santhanam, 2021). Recent developments in generative AI further extend these capabilities. Generative models demonstrate human-like reasoning, decision-making, and adaptive task execution—qualities essential for complex knowledge work (Banh & Strobel, 2023; Feuerriegel et al., 2024). When embedded in agentic IS artifacts, these models allow for the automation of entire roles, not merely isolated tasks. Building on this convergence, Ye et al. (2023) introduced the concept of agentic process automation (APA), which integrates generative AI agents into automated processes at points where dynamic decision-making is required. However, APA remains confined to predefined process points and lacks the interactive, decentralized coordination needed for more complex collaborative work. This limitation points to the need for MAS—a concept central to the agentic IS stream—where multiple AI-based agents collaborate within a shared environment. MAS, when populated by autonomous and intelligent agents, constitutes what is now termed MAAI, widely regarded as the next frontier in the automation of knowledge work (Guo et al., 2024; Yee, Chui, & Roberts, 2024).

MAAI integrates foundational elements from the three major research streams mentioned beforehand—process automation, agentic IS, and AI—resulting in MAAI systems capable of automating complex knowledge work processes previously considered inautomatable. Process automation contributes a focus on structured, process-oriented logic essential for and firmly anchored in digital economies. Agentic IS introduces the concept of autonomy of multiple independent entities, emphasizing agents capable of perceiving their environment and cooperating toward shared goals. AI, in turn, equips these agents with cognitive capabilities, enabling them to interpret context, make decisions, and adaptively act. Together, these streams form the conceptual foundation of MAAI. However, despite its transformative potential, the IS field still lacks a clear and comprehensive conceptualization of MAAI—one that distinguishes MAAI from conventional automation technologies such as RPA, accounts for its capacity to perform dynamic knowledge work, and positions it within existing IS streams, particularly around agentic IS. The following section builds on this foundation and introduces a conceptual framework that characterizes MAAI not as a rigid automation tool, but as a new class of intelligent systems.

3 Multi-Agent AI

To understand MAAI and its emerging role in automating complex knowledge work, we must move beyond the limitations of traditional process automation and its reliance on rigid, predefined workflows. MAAI marks a departure from such approaches by introducing a fundamentally different class of intelligent systems. To clarify what constitutes this new paradigm, we first provide a definition that captures the essential characteristics of MAAI:

Definition

Multi-Agent AI (MAAI) *refers to a system architecture composed of multiple autonomous or semi-autonomous AI-based agents that collaboratively perceive, reason, and act to perform context-sensitive, cognitive tasks. These agents dynamically structure workflows, coordinate roles across heterogeneous entities, including humans, and interact through interfaces within socio-technical environments.*

While this definition outlines *what* MAAI is, it leaves open the question of *how* such systems operate in practice. To address this, we conceptualize MAAI as a layered system of five stacked components (Cs)—each encapsulating and wrapping AI capabilities in a Matrjoschka-like architecture (Figure 1). These components do not enforce a fixed process but enable AI-based agents to collaboratively perceive, process, and act their path forward, moving beyond traditional RPA approaches. Moreover, MAAI is not designed to execute a rigid, known workflow; they are designed to dynamically construct agent-integrated workflows as they interact with data, other agents, and their environment. This section presents the framework's components, focusing on its agent-based dynamics, foundational technologies, and interaction interfaces. To enhance the comprehensibility of the framework and future applicability of MAAI, we present an illustrative exemplary use case of an insurance company displayed in fig. 2. This example, situated in the domain of digital markets, demonstrates the framework's capability to respond autonomously to customer inquiries regarding policy coverage for vehicle damage. Specifically, the automation involves the framework's components to determine whether the car damage claim falls within the scope of a customer's insurance policy. The customer submits the inquiry through a digital platform, for example, a mobile insurance app, including a textual description and images of the damaged cars and scenery.

Component C1: Foundation Model (FM)

The base of the framework rests on FMs, which refer to a class of large-scale models, pre-trained on vast amounts of data, that serve as a versatile base for a variety of downstream tasks (Bommasani et al., 2021; Schneider et al., 2024). Typically based on deep neural networks like transformers, FMs capture complex data relationships and can be fine-tuned for specific tasks or used as-is across various applications. FMs are often LLMs or large multimodal models (LMMs), such as GPT-x or Gemini. While LLMs solely provide natural language processing, LMMs also extend the ability to process further data modalities, such as vision or audio data. Yet, both unfold the computational capabilities essential for complex decision-making, reasoning and dealing with unstructured data (Lu et al., 2023; Paranjape et al., 2023; Schick et al., 2023). This makes them highly flexible and versatile for different tasks; thus, they act as a primary component of an AI-based agent and source of processing power within the framework both for taking in and generating unstructured data and for reasoning. To put FMs in the perspective of IS research, Schneider et al. (2024) introduce three key features that are of central importance to our framework: emergent capabilities, homogenization, and prompt sensitivity.

Emergent capabilities refer to skills that FMs develop during training, such as in-context learning (Min et al., 2022). This allows the model to perform new tasks, like solving mathematical problems or analogical reasoning, without explicit training for those tasks (Brown et al., 2020; Webb, Holyoak, & Lu, 2023). These capabilities are more pronounced in larger models like GPT-4 (Kaplan et al., 2020), and enable FMs to perform in various downstream tasks through prompt engineering, reducing the need for fine-tuning. However, the unpredictability of emergent behavior also introduces uncertainty, making it harder to foresee how the models will respond to slight changes in input. Homogenization addresses the concentration of a few dominant FMs, often developed by a few organizations using limited large datasets (Bommasani et al., 2021). This trend is driven by the high costs of training these models and the dominance of major companies with proprietary data. Homogenization has advantages, such as making AI development cheaper and faster by allowing advances in FMs to be inherited by downstream systems (Liu et al., 2023). However, it also centralizes power and raises concerns about dependency, ethical risks, and algorithmic monoculture, where biases and flaws from FMs are propagated across many AI applications (Fishman & Hancox-Li, 2022; Kleinberg & Raghavan, 2021). Prompt sensitivity refers to how the specific structure of input can significantly influence the behavior of FMs. The ability to guide and fine-tune model outputs using prompts is crucial for controlling and customizing

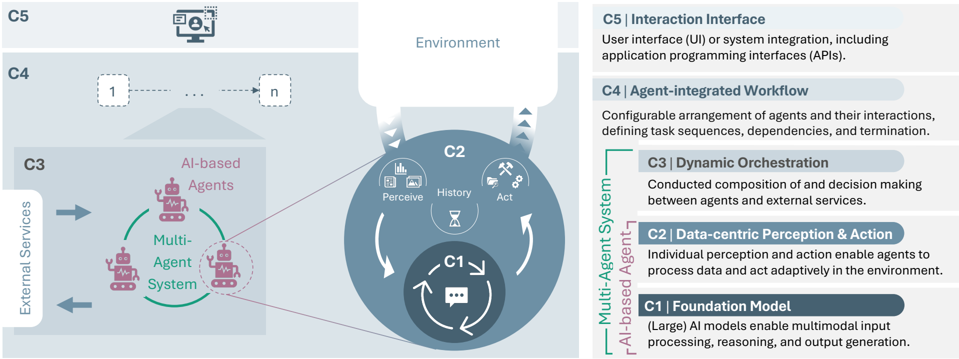

Fig. 1 Multi-Agent AI is characterized by a framework by a framework composed of a multi-agent-system (MAS) structured along five interrelated components. At its core, diverse Foundation Models (C1), such as large language or multimodal models, equip agents with the functionality to process multimodal inputs, integrate background knowledge, perform reasoning, and generate coherent outputs. The Data-centric Perception & Action (C2) component enables agents to gather, contextualize, and act upon information from their environment. Dynamic Orchestration (C3) manages the coordination and decision-making among multiple agents and external services, allowing flexible and adaptive task allocation. The Agent-integrated Workflow (C4) defines the sequence, interdependence, and logic of agent activities. Finally, the Interaction Interface (C5) connects the MAS to its users or external systems—either via user interfaces (UI) for human-AI collaboration or through application programming interfaces (APIs) for integration with other software systems.

agents. Small changes in prompts can yield vastly different outcomes, making prompt engineering—optimizing prompts for desired results—an essential technique (Schmidt, Spencer-Smith, Fu, & White, 2023). This also highlights challenges, such as adversarial attacks (jailbreaks) that exploit prompt vulnerabilities (A. Wei, Haghtalab, & Steinhardt, 2023).

In our exemplary insurance use case, FMs are crucial for handling the increasing complexity and data volume. Their scalability and generalization capabilities enable them to manage large and diverse inquiries—ranging from simple coverage questions to complex, multimodal inquiries—while drawing on a broad, pre-trained knowledge base. FMs, particularly LMMs, can seamlessly process and integrate various data types (e.g., textual claims or images of car damage) within a unified system, eliminating the need for different specialized AI models. Additionally, their ability to be fine-tuned allows them to adapt to the insurance particularities, ensuring long-term flexibility and relevance, such as standards and style of customer responses.

Component C2: Data-centric Perception & Action

Building upon FMs, AI-based agents comprise perception and action composing C2 to interact with their environment (Xi et al., 2023). In line with prior research in computer science, we define agents as entities that are modeled using concepts traditionally associated with human characteristics, such as autonomy, social ability, reactivity and pro-activeness (Goodwin, 1995; Wooldridge & Jennings, 1995). Rooted in MAS and

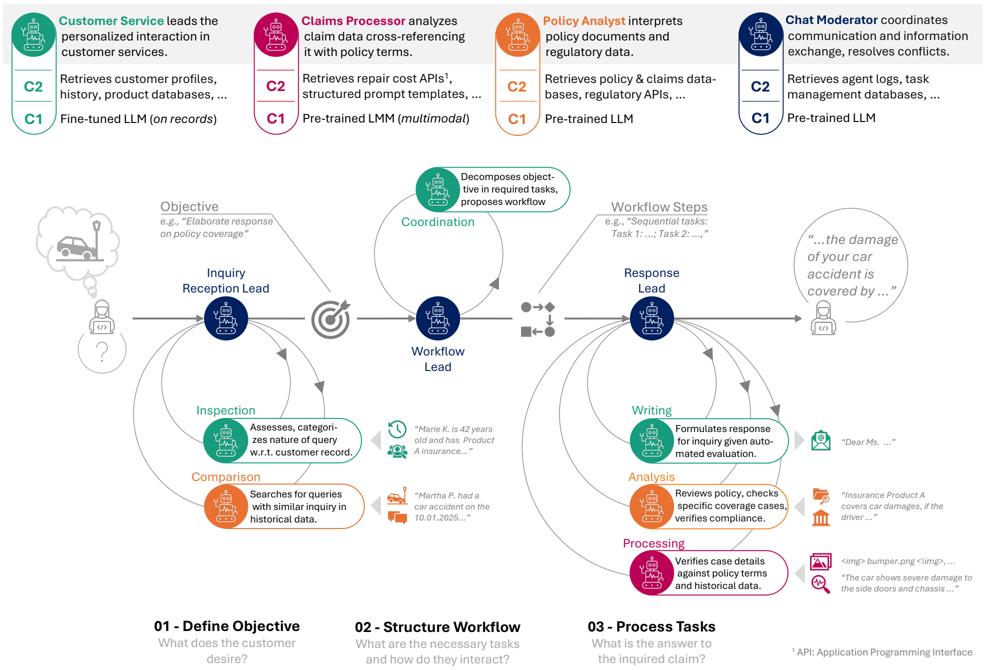

Fig. 2 An exemplary MAAI-based workflow in an insurance company setting. The illustration shows how multiple specialized AI-based agents, each leveraging a Foundation Model (C1) and a Data-centric Perception & Action component (C2), are dynamically orchestrated (C3) to form an Agent-integrated Workflow (C4). The workflow begins by defining the customer's objective—clarifying coverage for car damage—, structuring the necessary tasks and agent roles, and subsequently processing those tasks to deliver a comprehensive, policy-compliant response. Throughout this process, the Interaction Interface (C5) facilitates intuitive interaction, ensuring that human stakeholders can monitor or intervene as needed.

informed by agentic IS concepts, AI-based agents extend traditional process automation by incorporating human-centric capabilities like collaboration, decision-making, and adaptability. Consequently, each AI-based agent must incorporate mechanisms to perceive and act upon its environment, grounded in a data-centric approach (Jakubik, Vössing, Kühl, Walk, & Satzger, 2024). This approach acknowledges that while the underlying FM—its architecture and capacities—remain fixed, its effective adaptation and specialization are based on information in the form of data it receives and processes.

In data-centric AI, the primary idea for improving system performance and tailoring behaviors lies not in modifying the AI models' architecture but rather in refining, selecting, and structuring the data that the agent perceives and uses as context. Consequently, the perception component must sense and prepare diverse forms of data, optimizing them for the FM's internal representations. Although the FM's fundamental capabilities remain constant, the agent's skillfulness, domain relevance, and ability to handle evolving tasks are shaped by the data that it ingests and transforms into actionable insights. Since FMs store their learned representations in architectural parameters (weights) obtained from extensive, prior training, they only influence their performance by altering the data they process, both as input and through retrieval of external information. Perception in MAAI operates entirely on data: it transforms raw environmental inputs into structured, coherent formats that align with the FM's learned latent space of representations. This includes external data from sensors, documents, user inputs, application programming interfaces (APIs), and other sources, as well as internally managed context (e.g., previously generated text or prompt-engineered instructions). Through techniques like retrieval-augmented generation (RAG) (Gao et al., 2023; Lewis et al., 2020), structured templates (Schmidt et al., 2023; J. Wei et al., 2022), and APIs (Liang et al., 2024), the perception component curates and enhances the data delivered to the FM, effectively guiding the model's output without changing its core parameters. AI-based agents can access external databases or knowledge repositories through these mechanisms, retrieving relevant information dynamically to augment the context available to the FM. Here, the choice and quality of data—which documents to retrieve, what contexts to include, and how to structure the information—enforce distinct characteristics of the AI-based agent.

Once the FM processes the perceived data and produces an output, the action component interprets this output and executes actions in a data-mediated environment. Actions themselves are represented and performed through data: producing a textual response, generating a visual artifact, calling an API, or interacting with external systems are all rendered as data flows. This concept often involves leveraging specialized tools that enable the FM to perform specific functions, such as engaging in vocal conversations. By conceptualizing perception and action as data-centric processes—where even the agent's "touch" on its environment occurs through data-level transformations and interfaces—agents operate within and manipulate data representations. In addition to perception and action, C2 contains the capability to store and manage data, thereby increasing the contextual richness available to the FM. This stored context provides continuity across tasks, enabling AI-based agents to refine their decisions and responses based on historical and situational data.

Consider again the workflow of an insurance company handling a customer's car damage claim (Figure 2). Each specialized AI-based agent—be it the policy analyst, claims processor, or customer service agent—derives its expertise and context from both static and dynamic sources. Static elements such as predefined roles, agent profiles, and base prompts encode task-specific behavior and domain knowledge. Dynamically retrieved data, such as policy documents, customer records, or regulatory APIs, are then integrated during task execution to contextualize perception and guide action. For instance, the policy analyst can interpret insurance clauses by retrieving relevant documents, while the customer service agent adjusts tone and content by drawing on conversation history, customer records, and domain-specific prompts. After processing this curated input data, the FM's output directs the subsequent actions, enabling context-specific responses such as generating reports, sending messages, or integrating with external tools. These actions, too, remain data-driven: both perception and execution are shaped by what data can be accessed, structured, and operationalized in the agent's environment.

Component C3: Dynamic Orchestration

To effectively integrate the outputs and actions of various agents and maintain coherence within the agent-integrated workflow, AI-based agents must possess a distinct identity that grants them social standing (Bisk et al., 2020). This identity is essential for enabling social cooperation and competition within a MAS. The implementation of such a system comprises a registry of role-specific agents, each designed to be selected and deployed based on the demands of the task at hand (Pezeshkpour et al., 2024). This composition must be dynamically orchestrated in alignment with the specific requirements of the workflow. The orchestration should be sufficiently adaptable, allowing each workflow to autonomously determine its optimal configuration of agents. This ensures the orchestration can address the full complexity inherent in knowledge-intensive tasks. The orchestrated agents' capabilities are extended by connecting external services such as messaging applications or payment tools. In line with previous architectures like service-oriented computing, this integration allows orchestrated agents to execute actions that require interaction with external systems, broadening the scope of tasks they can handle (Calisti, Leymann, Dignum, Kowalczyk, & Unland, 2010; Tapia, Alonso, Zato, Gil, & de La Prieta, 2010; Ye et al., 2023).

In the context of MASs, the division of labor principle underpins the specialization of agents, where each agent, equipped with domain-specific expertise, is optimized for executing particular tasks. This specialization enhances the efficiency and precision of task execution, as it minimizes the overhead associated with task switching by decomposing complex workflows into discrete, manageable subtasks. Xi et al. (2023) delineate agent interactions into two primary paradigms: cooperative interaction for complementarity and adversarial interaction for strategic advancement. Cooperative interaction involves agents that continuously evaluate their counterparts' operational states and capabilities, engaging in collaborative task execution and knowledge exchange (Li, Hammoud, Itani, Khizbullin, & Ghanem, 2023). This interaction can be either disordered or ordered. In disordered cooperation, multiple agents operate with a high degree of autonomy, freely contributing their inputs and feedback within the system, which may result in misaligned objectives and potential disruptions in workflow orchestration (Mandi, Jain, & Song, 2024). Conversely, ordered cooperation adheres to predefined communication protocols, where agents sequentially process information, with downstream agents dependent on the outputs of upstream counterparts (Li et al., 2023; Talebirad & Nadiri, 2023). Dynamic orchestration mechanisms such as a moderator agent can be deployed to manage such interactions effectively to regulate communication flows and ensure adherence to structured review processes. In adversarial settings, agents dynamically reconfigure their strategies in response to real-time environmental feedback and the actions of competing agents (Abdelnabi, Gomaa, Sivaprasad, Schönherr, & Fritz, 2024). This adaptive behavior is crucial in adversarial MASs, where the objective is to optimize decision-making and enhance the quality of responses in high-stakes scenarios (Xi et al., 2023). AI-based agents may independently pursue optimization objectives, such as utility maximization, which aligns with principles of game theory (Mao et al., 2023). However, to maintain a structured competitive environment, deploying a moderator agent can facilitate the orchestration of interactions, ensuring that the competitive dynamics remain aligned with the overall system objectives.

In our insurance company use case, dynamic orchestration is essential for managing cooperative agent interactions. Cooperative interactions allow agents to independently contribute towards aligning on a common objective. For example, the policy analyst reviews the insurance policy for car insurance, while the claims processor verifies case details against historical data. A moderator agent is critical in this phase, regulating the dynamic information flow from both roles and ensuring alignment of agent outputs to produce a coherent and meaningful objective, thus maintaining the integrity of the MAAI. The cooperative scenario gains prominence in this use case when the objective is defined, allowing agents to engage in interactions that balance individual goals with overarching system objectives. In such contexts, the strategic oversight provided by a moderator is crucial for managing both cooperative and competitive dynamics, ensuring that the interplay between agents culminates in a cohesive and effective response to the customer inquiry.

Component C4: Agent-integrated Workflow

Given the inherent flexibility and complexity of many real-world tasks, they necessitate an agent-integrated workflow and dynamic decision-making process, which may be automatically created within MAAI (Ye et al., 2023). Rooted in the "information pull" paradigm, this component is conceptualized as a dynamic workflow, generated in direct response to emergent information needs (Cichocki, Helal, Rusinkiewicz, & Woelk, 1998). In the context of MAAI, this dynamic workflow is operationalized through the orchestration of AI-based agents, which execute tasks guided by abstract design patterns. Van der Aalst, ter Hofstede, Kiepuszewski, and Barros (2003) introduced a spectrum of design patterns, ranging from elementary constructs to intricate abstractions. The simpler design patterns are generally aligned with sequential or parallel workflows, facilitating the pre-definition of workflow steps and enabling task-oriented deployment through high-level directives (Xi et al., 2023). This approach facilitates the orchestration to establish all requisite tasks in advance, allowing AI-based agents to execute them autonomously. Each task's outcome is then sequentially forwarded to the subsequent task. Conversely, more intricate workflows exploit the full intellectual potential of MASs through sophisticated orchestration. These workflows incorporate iterative loops, necessitating innovation-oriented deployment or even lifecycle-oriented deployment, enabling agents to continuously interact with an open and unknown environment (Xi et al., 2023).

In our example, the customer's inquiry is addressed through a three-stage process: initially, the overall objective is identified, followed by the structuring of the workflow, and culminating in the processing of the tasks, all within the collaborative framework facilitated by the dynamic orchestration of the MAS. Each stage supports dynamic orchestration and decision-making, allowing for adaptive and responsive interactions throughout the workflow. A policy analyst agent might need to query databases an unknown number of times to gather sufficient information about a case. In such a complex scenario, the agent-integrated workflow component showcases innovation-oriented deployment, iteratively adjusting tasks as the process evolves. This method enhances flexibility and adaptability within workflow automation, accommodating changes and enabling dynamic decision-making processes that emulate human-like deliberations rather than adhering to rigid rule-based protocols.

Component C5: Interaction Interface

The interaction interface serves as the entry point for both human and system-level engagement with MAAI. It encompasses not only user interfaces (UIs) but also APIs, enabling integration with external systems and organizational platforms. While the previous components enable AI-based agent coordination and agent-integrated workflow generation, the interaction interface ensures that these processes are accessible, traceable, and actionable. As such, the UI constitutes the critical socio-technical junction where the system becomes actionable in organizational settings. Effective user interfaces are also essential for enabling human oversight and intervention, especially in complex, automated workflows (Sterz et al., 2024). Drawing on IS perspectives (Baird & Maruping, 2021; Schuetz & Venkatesh, 2020), this component plays a dual role: supporting human-AI collaboration through explainable, adaptive, and role-specific interfaces, and facilitating machine-to-machine communication through structured endpoints. In organizational settings, such interfaces must balance transparency with usability—offering dashboards, conversational agents, or monitoring tools for internal users, and streamlined interaction channels for external systems. This goes beyond the necessity of having any interaction interface—it raises design requirements related to explainability, adaptive feedback, and trust-building mechanisms in human-AI collaboration. Designing an appropriate interaction interface for MAAI thus depends on the system's purpose and user roles. For end-users, chatbot interfaces or decision dashboards may suffice to deliver responses and gather input. For internal users (e.g., employees monitoring agent-integrated workflows), the UI must support traceability of agent interactions, intervention capabilities, and contextual awareness. In both cases, prior work in IS suggests the importance of adaptive, stateful interfaces and role-specific views to manage complexity and promote trust (Gimpel, Graf-Drasch, Laubacher, & Wöhl, 2020; Gnewuch, Morana, & Maedche, 2017).

In the insurance company use case, the interaction interface connects customers, employees, and the MAAI. A chatbot UI allows customers to submit claims and receive updates, while internal APIs dynamically feed outputs from policy analysts or claims processors into case management systems. These interaction interfaces ensure that the system remains comprehensible, responsive, and compatible with existing workflows.

4 Challenges and Emerging Pathways for the IS Community

The advent of MAAI represents a shift from conventional, narrowly defined automation toward more sophisticated, agent-based AI systems capable of addressing intricate problem-solving and decision-making tasks. FMs, especially those found in generative AI, represent a transformative potential for automating a wide range of knowledge work. For the IS community, this development requires not only theoretical reflection but also actionable guidance in research avenues spanning technical capabilities and

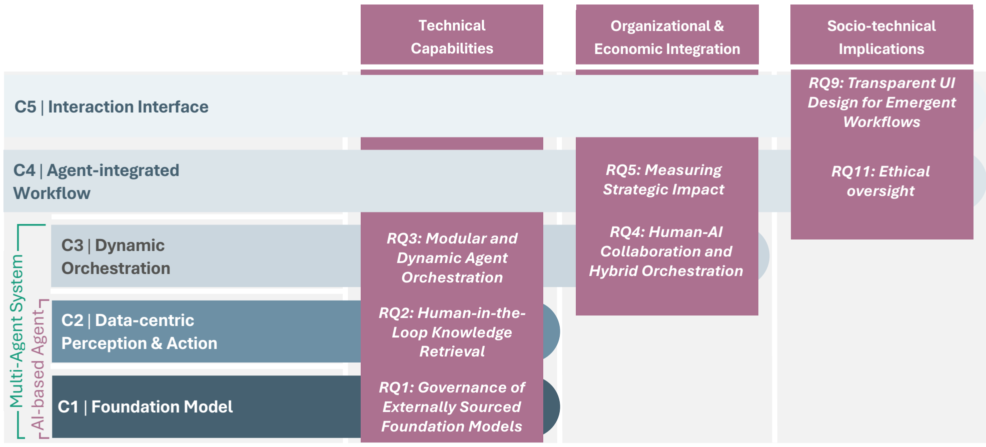

Fig. 3 The figure maps the five-component MAAI framework—comprising FM, data-centric perception & action, dynamic orchestration, agent-integrated workflow, and interaction interface—to exemplary research questions from the three proposed analytical pathways: technical capabilities, organizational and economic integration, and socio-technical implications.

organizational-economic integration alongside socio-technical implications. The conceptual framework presented in this work offers such guidance. It is intended to inform and structure developing or analyzing MAAI. Each component addresses a core architectural layer and provides a scaffold to formulate requirements, select technologies, define agent roles, and assess socio-technical implications. Researchers can use the framework to compare system architectures, identify gaps in existing solutions, and formulate research questions grounded in component-specific challenges (cf. Figure 3). Practitioners may use it as a reference for implementing modular, agent-based systems in real-world settings.

4.1 *Engineer* Technical Capabilities (Components C1-C3)

MAAI's potential to support employees and alleviate burdensome workloads introduces significant engineering challenges, particularly in designing coherent systems that enable the dynamic orchestration (C3) of agents within context-sensitive workflows. A key issue is that FMs (C1)—often developed externally and accessed via APIs—limit organizational control over model evolution (e.g., domain-specific tuning (Ray, 2023)), functionality, and propagated bias (Ji et al., 2023). In MAAI, faulty outputs from one agent can influence others, compounding errors across the system. This raises the need for processes that support ongoing alignment between model capabilities, organizational needs, and customer-facing outcomes. To address these challenges, applying design science research approaches (Tuunanen, Winter, & vom Brocke, 2024) offers a structured foundation for deriving modular, adaptive design patterns for MAAI. Existing frameworks remain largely conceptual (Händler, 2023; J. Zhou et al., 2024) and lack prescriptive guidance for implementing agent-based AI systems in practical, value-generating contexts (Chen et al., 2024; Crawford et al., 2024; Durante et al., 2024). Researchers must develop reference architectures, decision models, and modular software components that operationalize the orchestration of heterogeneous agents (C3), informed by domain-specific and organizational constraints. Beyond technical orchestration, the effectiveness of data-centric perception and action (C2) hinges on how agents access and interpret relevant information. Current data retreival techniques rely heavily on static documentation, which limits responsiveness to nuanced, real-world practices. There is a need to incorporate human-generated insights—capturing how tasks are performed in practice—into retrieval mechanisms, thereby increasing contextual sensitivity and enhancing agent performance in dynamic environments. These challenges give rise to the following research questions:

- RQ1 **Governance of Externally Sourced Foundation Models (C1)** Given that foundation models are often externally developed and integrated via APIs, leaving development teams with limited control over their evolution, how can organizations establish continuous development and deployment processes that align model capabilities with domain-specific requirements and mitigate risks related to opacity, bias, and functional drift?
- RQ2 **Human-in-the-Loop Knowledge Retrieval (C2)** How can human-generated insights and tacit expertise be systematically integrated into knowledge retrieval mechanisms to enhance contextual relevance and move beyond static sources such as documentation or wikis?
- RQ3 **Modular and Dynamic Agent Orchestration (C3)** How can design principles be applied to enable modular and dynamic orchestration of heterogeneous AI-based agents within MAAI systems, allowing workflows to be adaptively structured and refined in real-time as tasks evolve?

4.2 *Structure* Organizational and Economic Integration (Components C3-C4)

From a managerial standpoint, the integration of MAAI into enterprise workflows raises fundamental questions about the future of knowledge work and the evolving role of human labor. While AI-based agents are increasingly capable of autonomously executing complex, multi-step processes, human involvement currently remain essential—not only for oversight and exception handling (Yee et al., 2024), but also as a meaningful component within agentic ecosystems. Recent research suggests that humans can function as "human agents" within MAAI (Guo et al., 2024; Song et al., 2024), collaborating with AI-based agents to enrich decision-making, ensure accountability (cf. European Union AI Act), and provide situational awareness in dynamic environments. This shift toward hybrid orchestration (C3) necessitates the development of mechanisms that dynamically allocate tasks to both human and AI-based agents based on context-sensitive evaluations of skill, expertise, and value contribution. Beyond static role definitions, such integration requires flexible orchestration logic capable of adapting workflows to real-time demands, ensuring that human input is solicited when it is most impactful (Hemmer, Schemmer, Kühl, Vössing, &

Satzger, 2024). Technical evaluations have already found that multiple AI agents working together do not always outperform single agents, highlighting the need for nuanced analysis (Cemri et al., 2025). The practical realization of this vision also entails significant change management: determining which workflows are best suited for full automation, which require human-AI collaboration, and how to position humans effectively within the broader system architecture. From an organizational perspective, the deployment of MAAI is expected to create new forms of work and alter the composition of existing roles. As agent-based automation is embedded into workflows (C4), organizations must assess the emergence of new professional categories and consider how current fields evolve. Strategic implementation thus requires not only technical integration but also workforce planning, training, and governance mechanisms to support long-term adaptation. Finally, the economic viability of MAAI also depends on resource efficiency. Not all agents require large FMs; orchestration mechanisms (C3) must assign tasks based on the complexity and data sensitivity of the task, leveraging lightweight or specialized models where feasible. Techniques such as model parallelization, quantization, or knowledge distillation can further reduce the computational burden, making MAAI solutions more scalable and cost-effective (Zhu et al., 2024). In light of these developments, we identify the following research questions:

- RQ4 **Human-AI Collaboration and Hybrid Orchestration (C3):** How can humans be integrated as active agents within MAAI, and how can orchestration mechanisms dynamically allocate human and AI-based agents to tasks based on the comparative value of human skills?
- RQ5 **Measuring Strategic Impact (C4):** What metrics and evaluation frameworks can be developed to assess the strategic and financial benefits of integrating orchestrated AI-based agents into real-world organizational workflows? And how do we measure the performance of a MAAI in a business process—e.g., by aggregate outcome, by human satisfaction, or by emergent properties?
- RQ6 **Role Transformation and Emerging Professions (C3, C4):** Which tasks and workflows are most impacted by the deployment of MAAI, and what new professional roles (e.g., operations, orchestration design) are likely to emerge as organizations adapt to MAAI?
- RQ7 **Resource Efficiency and Economic Scalability (C3):** How can orchestration strategies be optimized to ensure resource-efficient MAAI deployments, balancing task complexity with the use of lightweight models, parallelization, and modular architectures?

4.3 *Study* Socio-technical Implications (Components C4-C5)

The societal perspectives of MAAI span ethical, labor, sustainability, and governance concerns. As AI-based agents increasingly assume complex decision-making tasks, the labor market may see both displacement and transformation of roles. While automation can democratize access to expert-like capabilities, it also risks downward pressure on wages and the erosion of returns to general education (Acemoglu & Restrepo, 2020; Agrawal, Gans, & Goldfarb, 2023; Dwivedi et al., 2023). These dynamics call for proactive workforce planning and targeted interventions to mitigate inequality. On the ethical front, preserving human responsibility becomes increasingly difficult as agent interactions grow more autonomous and interdependent. Faulty outputs from one agent may propagate through others, compounding risks. Transparent evaluation and governance structures are essential to avoid overreliance and preserve accountability (Deutscher Ethikrat, 2023; Ray, 2023; Schemmer et al., 2023). MAAI particularly raises questions about fairness and accountability in decision-making workflows. Outputs can be evaluated at both final and intermediate stages, yet it remains unclear whether fairness principles must apply uniformly—especially when not all outputs are user-facing. This prompts a critical inquiry into whether interactions among AI-based agents themselves must adhere to ethical standards reflective of human stakeholder norms. From a socio-technical angle, UIs (C5) play a central role in making MAAI interpretable. In agent integrated workflows (C4), where tasks and agent configurations evolve on the fly, interfaces must provide transparency without cognitive overload. Long-form textual outputs of agent dialogues may obscure rather than support informed oversight, complicating appropriate reliance (Schemmer et al., 2023) and hence confident adoption (Ray, 2023). Research questions for the third research avenue include:

- RQ8 **Multi-level Fairness in MAAI Workflows (C4)**: To what extent should principles of algorithmic fairness be applied uniformly—particularly when certain outputs may not directly interface with human stakeholders?
- RQ9 **Transparent UI Design for Emergent Workflows (C5)**: How can interaction interfaces be designed to transparently present the evolving outputs and decisions of MAAI systems, especially in workflows where tasks and agent roles are dynamically generated, and how can such interfaces promote appropriate reliance without overwhelming users with excessive or unintelligible information?
- RQ10 **Labor Market Dynamics and Human Integration via Orchestration (C3, C4)**: How will the widespread adoption of MAAI reshape labor markets, particularly as human roles become dynamically integrated into orchestrated workflows, and what policy interventions or training programs are needed to address resulting changes in employment levels, wage structures, and skill demands?
- RQ11 **Ethical oversight (C4)**: How can agent-integrated workflows be designed to maintain transparency, draw boundaries of human responsibility, and prevent the undermining of accountability?

5 Conclusion and Research Directions

Multi-Agent AI (MAAI) marks a foundational shift in the automation of knowledge work—moving beyond rigid task execution toward adaptive, agent-based collaboration grounded in perception, reasoning, and dynamic orchestration. Rooted in the traditions of process automation, agentic information systems, and artificial intelligence, MAAI responds to the increasing complexity of knowledge work illustrated in our opening scenario: work that spans organizational boundaries and demands contextual, expert-driven decision-making. To structure this emerging domain, we proposed a layered five-component framework: Foundation model (C1), data-centric perception and action (C2), dynamic orchestration (C3), agent-integrated workflow (C4), and user interface (C5). This Matrjoschka-like architecture not only explains how MAAI systems function, but also structures the debate—unpacking the technical foundations, interaction mechanisms, and human interfaces that define MAAI's unique socio-technical dynamics. The framework delineates three critical research avenues. Technically, it calls for new approaches to agent modularity, orchestration, and perception in systems that increasingly rely on opaque, externally developed models. Organizationally, it prompts inquiry into hybrid human-AI workflows, changing labor roles, and value distribution. Socio-technically, it highlights the need for accountable, transparent systems capable of supporting trust, fairness, and responsible oversight in multi-task settings.

By engaging these challenges, the IS community can shape MAAI not just as a technical evolution, but as a transformative paradigm for how intelligence, coordination, and agency are distributed in digital systems—grounding future innovation of MAAI in both conceptual clarity and societal relevance.

References

Abdelnabi, S., Gomaa, A., Sivaprasad, S., Schönherr, L., Fritz, M. (2024). Cooperation, competition, and maliciousness: LLM-stakeholders interactive negotiation. *Advances in Neural Information Processing Systems* (Vol. 37, pp. 83548–83599). Vancouver, Canada: Curran Associates, Inc.

Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy*, *128*(6), 2188–2244, https://doi.org/10.1086/705716

Acharya, D.B., Kuppan, K., Divya, B. (2025). Agentic AI: Autonomous intelligence for complex goals - a comprehensive survey. *IEEE Access*, *13*, 18912-18936, https://doi.org/10.1109/ACCESS.2025.3532853

Agrawal, A., Gans, J.S., Goldfarb, A. (2023). Do we want less automation? *Science*, *381*(6654), 155–158, https://doi.org/10.1126/science.adh9429

Anonymous (2025). *Agentic information systems.* (Working Paper)

Autor, D.H., Levy, F., Murnane, R.J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, *118*(4), 1279–1333, https://doi.org/10.1162/003355303322552801

Baird, A., & Maruping, L.M. (2021). The next generation of research on is use: A theoretical framework of delegation to and from agentic is artifacts. *MIS Quarterly*, *45*(1), 315–341, https://doi.org/10.25300/MISQ/2021/15882

Banh, L., & Strobel, G. (2023). Generative artifcial intelligence. *Electronic Markets*, *33*(63), 1–17, https://doi.org/10.1007/s12525-023-00680-1

Berente, N., Gu, B., Recker, J., Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, *45*(3), 1433-1450, https://doi.org/10.25300/MISQ/2021/16274

Bisk, Y., Holtzman, A., Thomason, J., Andreas, J., Bengio, Y., Chai, J., … Turian, J. (2020). Experience grounds language. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 8718–8735). online: Association for Computational Linguistics.

Bock, A.C., & Frank, U. (2021). Low-code platform. *Business & Information Systems Engineering*, *63*(6), 733–740, https://doi.org/10.1007/s12599-021-00726-8

Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., Arx, S.v., … Liang, P. (2021). *On the opportunities and risks of foundation models.* (*ArXiv preprint ArXiv:2108.07258*)

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., … Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems* (Vol. 33, pp. 1877–1901). Curran Associates, Inc.

Calisti, M., Leymann, F., Dignum, F.P., Kowalczyk, R., Unland, R. (2010). Service-oriented architecture and (multi-)agent systems technology. *Dagstuhl Seminar Proceedings* (Vol. 10021, pp. 1–16). Wadern, Germany.

Cemri, M., Pan, M.Z., Yang, S., Agrawal, L.A., Chopra, B., Tiwari, R., … Stoica, I. (2025). *Why do multi-agent llm systems fail?* (*ArXiv preprint ArXiv:2503.13657*)

Chen, W., You, Z., Li, R., Guan, Y., Qian, C., Zhao, C., … Sun, M. (2024). *Internet of agents: Weaving a web of heterogeneous agents for collaborative intelligence.* (*ArXiv preprint ArXiv:2407.07061*)

Cichocki, A., Helal, A., Rusinkiewicz, M., Woelk, D. (1998). *Workflow and process automation: Concepts and technology.* London, UK: Kluwer Academic Publishers.

Crawford, N., Duffy, E.B., Evazzade, I., Foehr, T., Robbins, G., Saha, D.K., ... Ziolkowski, M. (2024). *BMW agents–a framework for task automation through multi-agent collaboration.* (*ArXiv preprint ArXiv:2406.20041*)

Davenport, T.H., Jarvenpaa, S., Beers, M. (1996). Improving knowledge work processes. *Sloan Management Review*, *37*, 53–66, Retrieved from https://sloanreview.mit.edu/article/improving-knowledge-work-processes/

Davenport, T.H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know* (Vol. 5). Boston, MA, USA: Harvard Business School Press.

Deutscher Ethikrat (2023). *Mensch und Maschine - Herausforderungen durch Künstliche Intelligenz.* (Stellungnahme)

Dorri, A., Kanhere, S.S., Jurdak, R. (2018). Multi-agent systems: A survey. *IEEE Access*, *6*, 28573–28593, https://doi.org/10.1109/ACCESS.2018.2831228

Drucker, P.F. (1999). Knowledge-worker productivity: The biggest challenge. *California Management Review*, *41*(2), 79–94, https://doi.org/10.2307/4116598

Durante, Z., Sarkar, B., Gong, R., Taori, R., Noda, Y., Tang, P., ... Huang, Q. (2024). *An interactive agent foundation model.* (*ArXiv preprint ArXiv:2402.05929*)

Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kar, A.K., ... Wright, R. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642, https://doi.org/10.1016/j.ijinfomgt.2023.102642

Engel, C., Ebel, P., Leimeister, J.M. (2022). Cognitive automation. *Electronic Markets*, *32*(1), 339–350, https://doi.org/10.1007/s12525-021-00519-7

Ferber, J., & Weiss, G. (1999). *Multi-agent systems: an introduction to distributed artificial intelligence* (Vol. 1). Addison-wesley Reading.

Feuerriegel, S., Hartmann, J., Janiesch, C., Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, *66*(1), 111–126, https://doi.org/

10.1007/s12599-023-00834-7

Fishman, N., & Hancox-Li, L. (2022). Should attention be all we need? The epistemic and ethical implications of unification in machine learning. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1516–1527). Seoul, South Korea: Association for Computing Machinery.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... Wang, H. (2023). *Retrieval-augmented generation for large language models: A survey.* (*ArXiv preprint ArXiv:2312.10997*)

Gimpel, H., Graf-Drasch, V., Laubacher, R.J., Wöhl, M. (2020). Facilitating like darwin: Supporting cross-fertilisation in crowdsourcing. *Decision Support Systems*, *132*, 113282, https://doi.org/10.1016/j.dss.2020.113282

Gnewuch, U., Morana, S., Maedche, A. (2017). Towards designing cooperative and social conversational agents for customer service. *Proceedings of the International Conference on Information Systems (ICIS).* Seoul, South Korea: Association for Information Systems.

Goodwin, R. (1995). Formalizing properties of agents. *Journal of Logic and Computation*, *5*(6), 763–781, https://doi.org/10.1093/logcom/5.6.763

Guo, H., Mu, C., Chen, Y., Shen, C., Hu, S., Wang, Z. (2024). *Multi-agent, human-agent and beyond: A survey on cooperation in social dilemmas.* (*ArXiv preprint ArXiv:2402.17270*)

Göldi, A., & Rietsche, R. (2024). Making sense of large language model-based AI agents. *Proceedings of the International Conference on Information Systems (ICIS).* Bangkok, Thailand: Association for Information Systems. Retrieved from https://aisel.aisnet.org/icis2024/aiinbus/aiinbus/16/

Haase, J., Kremser, W., Leopold, H., Mendling, J., Onnasch, L., Plattfaut, R. (2024). Interdisciplinary directions for researching the effects of robotic process automation and large language models on business processes. *Communications of the Association for Information Systems*, *54*(1), 579–604, https://doi.org/10.17705/1CAIS.05421

Händler, T. (2023). *Balancing autonomy and alignment: a multi-dimensional taxonomy for autonomous LLM-powered multi-agent architectures.* (*ArXiv preprint ArXiv:2310.03659*)

Heerwagen, J.H., Kampschroer, K., Powell, K.M., Loftness, V. (2004). Collaborative knowledge work environments. *Building Research & Information*, *32*(6), 510–528, https://doi.org/10.1080/09613210412331313025

Hemmer, P., Schemmer, M., Kühl, N., Vössing, M., Satzger, G. (2024). *Complementarity in human-ai collaboration: Concept, sources, and evidence.* (*ArXiv preprint ArXiv: 2404.00029*)

Hofmann, P., Samp, C., Urbach, N. (2020). Robotic process automation. *Electronic Markets*, *30*(1), 99–106, https://doi.org/10.1007/s12525-019-00365-8

Hughes, L., Dwivedi, Y.K., Malik, T., Shawosh, M., Albashrawi, M.A., Jeon, I., … Walton, P. (2025). AI agents and agentic systems: A multi-expert analysis. *Journal of Computer Information Systems*, 1–29, https://doi.org/10.1080/08874417.2025.2483832

IEEE guide for terms and concepts in intelligent process automation. (2017). *IEEE Std 2755-2017*, 1-16, https://doi.org/10.1109/IEEESTD.2017.8070671

Jakubik, J., Vössing, M., Kühl, N., Walk, J., Satzger, G. (2024). Data-centric artificial intelligence. *Business & Information Systems Engineering*, *66*, 507–515, https://doi.org/10.1007/s12599-024-00857-8

Janiesch, C., Zschech, P., Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695, https://doi.org/10.1007/s12525-021-00475-2

Janz, B., Conquitt, J., Noe, R. (1997). Knowledge worker team effectiveness: The role of autonomy, interdependence, team development, and contextual support. *Personnel Psychology*, *50*(4), 877–904, https://doi.org/10.1111/j.1744-6570.1997.tb01486.x

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., … Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, *55*(12), 1–38, https://doi.org/10.1145/3571730

Jimenez-Ramirez, A., Reijers, H.A., Barba, I., Del Valle, C. (2019). A method to improve the early stages of the robotic process automation lifecycle. P. Giorgini et al. (Eds.), *Advanced Information Systems Engineering* (Vol. 11483, pp. 446–461). Cham, Germany: Springer International Publishing.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., ... Amodei, D. (2020). *Scaling laws for neural language models.* (*ArXiv preprint ArXiv:2001.08361*)

Kleinberg, J., & Raghavan, M. (2021). Algorithmic monoculture and social welfare. *Proceedings of the National Academy of Sciences*, *118*(22), 1–7, https://doi.org/10.1073/pnas.2018340118

Kuehl, N., Schemmer, M., Goutier, M., Satzger, G. (2022). Artificial intelligence and machine learning. *Electronic Markets*, *32*(4), 2235–2244, https://doi.org/10.1007/s12525-022-00598-0

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Conference on Neural Information Processing Systems* (Vol. 33, pp. 9459–9474). Vancouver, Canada.

Li, G., Hammoud, H.A.A.K., Itani, H., Khizbullin, D., Ghanem, B. (2023). Camel: Communicative agents for “mind exploration of large language model society. *Conference on Neural Information Processing Systems* (pp. 51991–52008). New Orleans, Louisiana, USA.

Liang, Y., Wu, C., Song, T., Wu, W., Xia, Y., Liu, Y., ... Duan, N. (2024). TaskMatrix.AI: Completing tasks by connecting foundation models with millions of APIs. *Intelligent Computing*, *3*, 0063, https://doi.org/10.34133/icomputing.0063

Lins, S., Pandl, K.D., Teigeler, H., Thiebes, S., Bayer, C., Sunyaev, A. (2021). Artificial intelligence as a service. *Business & Information Systems Engineering*, *63*(4), 441–456, https://doi.org/10.1007/s12599-021-00708-w

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, *55*(9), 1–35, https://doi.org/10.1145/3560815

Lu, P., Peng, B., Cheng, H., Galley, M., Chang, K.-W., Wu, Y.N., ... Gao, J. (2023). Chameleon: Plug-and-play compositional reasoning with large language models. *Conference on Neural Information Processing Systems* (Vol. 36, pp.

43447–43478). New Orleans, Louisiana, USA: Curran Associates, Inc.

Mandi, Z., Jain, S., Song, S. (2024). RoCo: Dialectic multi-robot collaboration with large language models. *2024 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 286–299). IEEE.

Mao, S., Cai, Y., Xia, Y., Wu, W., Wang, X., Wang, F., ... Wei, F. (2023). *ALYMPICS: LLM agents meet game theory – exploring strategic decision-making with AI agents.* (*ArXiv preprint ArXiv:2311.03220*)

Min, S., Lyu, X., Holtzman, A., Artetxe, M., Lewis, M., Hajishirzi, H., Zettlemoyer, L. (2022). Rethinking the role of demonstrations: What makes in-context learning work? Y. Goldberg et al. (Eds.), *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 11048–11064). Stroudsburg, PA, USA: Association for Computational Linguistics.

Paranjape, B., Lundberg, S., Singh, S., Hajishirzi, H., Zettlemoyer, L., Ribeiro, M.T. (2023). *Art: Automatic multi-step reasoning and tool-use for large language models.* (*ArXiv preprint ArXiv:2303.09014*)

Pezeshkpour, P., Kandogan, E., Bhutani, N., Rahman, S., Mitchell, T., Hruschka, E. (2024). *Reasoning capacity in multi-agent systems: Limitations, challenges and human-centered solutions.* (*ArXiv preprint ArXiv:2402.01108*)

Pyöriä, P. (2005). The concept of knowledge work revisited. *Journal of Knowledge Management*, *9*(3), 116–127, https://doi.org/10.1108/13673270510602818

Ray, P.P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, *3*, 121–154, https://doi.org/10.1016/j.iotcps.2023.04.003

Russell, S., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (3rd ed.). Boston, MA, USA: Pearson.

Schemmer, M., Kuehl, N., Benz, C., Bartos, A., Satzger, G. (2023). Appropriate reliance on ai advice: Conceptualization and the effect of explanations. *International Conference on Intelligent User Interfaces* (pp. 410–422). Sydney, NSW, Australia: Association for Computing Machinery.

Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., ... Scialom, T. (2023). *Toolformer: Language models can teach themselves to use tools.* (*ArXiv preprint ArXiv:2302.04761*)

Schmidt, D.C., Spencer-Smith, J., Fu, Q., White, J. (2023). *Cataloging prompt patterns to enhance the discipline of prompt engineering.* Retrieved from https://www.dre.vanderbilt.edu/~schmidt/PDF/ADA_Europe_Position_Paper.pdf

Schneider, J., Meske, C., Kuss, P. (2024). Foundation models: A new paradigm for artificial intelligence. *Business & Information Systems Engineering*, *66*(2), 221–231, https://doi.org/10.1007/s12599-024-00851-0

Schuetz, S., & Venkatesh, V. (2020). The rise of human machines: How cognitive computing systems challenge assumptions of user-system interaction. *Journal of the Association for Information Systems*, *21*(2), 460–482, Retrieved from https://ssrn.com/abstract=3680306

Schultze, U. (2000). A confessional account of an ethnography about knowledge work. *MIS Quarterly*, *24*(1), 3–41, https://doi.org/10.2307/3250978

Song, T., Tan, Y., Zhu, Z., Feng, Y., Lee, Y.-C. (2024). *Multi-agents are social groups: Investigating social influence of multiple agents in human-agent interactions.* (*ArXiv preprint ArXiv:2411.04578*)

Sterz, S., Baum, K., Biewer, S., Hermanns, H., Lauber-Rönsberg, A., Meinel, P., Langer, M. (2024). On the quest for effectiveness in human oversight: Interdisciplinary perspectives. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 2495–2507). Rio de Janeiro, Brazil: ACM.

Svanberg, M., Li, W., Fleming, M., Goehring, B., Thompson, N. (2024). *Beyond AI exposure: Which tasks are cost-effective to automate with computer vision?* Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4700751 (SSRN Working Paper)

Talebirad, Y., & Nadiri, A. (2023). *Multi-agent collaboration: Harnessing the power of intelligent LLM agents.* (*ArXiv preprint ArXiv:2306.03314*)

Tapia, D.I., Alonso, R.S., Zato, C., Gil, O., de La Prieta, F. (2010). Analysis and design of a SOA-based multi-agent architecture. *Trends in Practical Applications of Agents and Multiagent Systems* (Vol. 71, pp. 183–190). Berlin, Germany: Springer.

Tuunanen, T., Winter, R., vom Brocke, J. (2024). Dealing with complexity in design science research: A methodology using design echelons. *MIS Quarterly*, *48*(2), 427–458, https://doi.org/10.25300/MISQ/2023/16700

Van der Aalst, W., ter Hofstede, A., Kiepuszewski, B., Barros, A.P. (2003). Workflow patterns. *Distributed and Parallel Databases*, *14*(1), 5–51, https://doi.org/10.1023/A:1022883727209

Webb, T., Holyoak, K.J., Lu, H. (2023). Emergent analogical reasoning in large language models. *Nature Human Behaviour*, *7*(9), 1526–1541, https://doi.org/10.1038/s41562-023-01659-w

Wei, A., Haghtalab, N., Steinhardt, J. (2023). Jailbroken: How does llm safety training fail? *Advances in Neural Information Processing Systems* (Vol. 36, pp. 80079–80110). Curran Associates, Inc.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., ... Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* (Vol. 35, pp. 24824–24837). Curran Associates, Inc.

Willcocks, L., Lacity, M., Craig, A. (2017). Robotic process automation: Strategic transformation lever for global business services? *Journal of Information Technology Teaching Cases*, *7*(1), 17–28, https://doi.org/10.1057/s41266-016-0016-9

Wooldridge, M. (2009). *An introduction to multiagent systems.* John Wiley & Sons.

Wooldridge, M., & Jennings, N.R. (1995). Intelligent agents: theory and practice. *The Knowledge Engineering Review*, *10*(2), 115–152, https://doi.org/10.1017/S0269888900008122

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., ... Gui, T. (2023). *The rise and potential of large language model based agents: A survey.* (*ArXiv preprint ArXiv:2309.07864*)

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., ... Gui, T. (2025). The rise and potential of large language model based agents: A survey. *Science China Information Sciences*, *68*(2), 121101, https://doi.org/10.1007/s11432-024-4222-0

Ye, Y., Cong, X., Tian, S., Cao, J., Wang, H., Qin, Y., ... Sun, M. (2023). *ProAgent: From robotic process automation to agentic process automation.* (*ArXiv preprint ArXiv:2311.10751*)

Yee, L., Chui, M., Roberts, R. (2024). *Why agents are the next frontier of generative AI.* Retrieved from https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/why%20agents%20are%20the%20next%20frontier%20of%20generative%20ai/why-agents-are-the-next-frontier-of-generative-ai.pdf?shouldIndex=false

Zhou, J., Lu, Q., Chen, J., Zhu, L., Xu, X., Xing, Z., Harrer, S. (2024). *A taxonomy of architecture options for foundation model-based agents: Analysis and decision model.* (*ArXiv preprint ArXiv:2408.02920*)

Zhou, W., Jiang, Y.E., Li, L., Wu, J., Wang, T., Qiu, S., ... Sachan, M. (2023). *Agents: An open-source framework for autonomous language agents.* (*ArXiv preprint ArXiv:2309.07870*)

Zhu, X.X., Xiong, Z., Wang, Y., Stewart, A.J., Heidler, K., Wang, Y., ... Shi, Y. (2024). *On the foundations of earth and climate foundation models.* (*ArXiv preprint ArXiv:2405.04285*)