



CATCHWORD

Agentic AI

From Autonomous Agents to Enterprise Agentic AI Systems

Mahei Manhai Li · Philipp Reinhard · Jan Marco Leimeister

Received: 10 July 2025 / Accepted: 29 April 2026
© The Author(s) 2026

Keywords Agentic AI · AI agents · Multi-agent systems (MAS) · Generative AI (GenAI) · Digital agency · IS delegation

1 Introduction

Artificial intelligence (AI) is shifting from simple automation (Janiesch et al. 2021) and content generation (Feuerriegel et al. 2024) towards solving complex goals across enterprise landscapes (Baird and Maruping 2021). Enabled by advances in generative AI (GenAI) and large language models (LLMs) (Feuerriegel et al. 2024; Schneider et al. 2024), tool learning (Qin et al. 2025; Qu et al. 2025), as well as large action models (LAMs) (Zhang et al. 2025), agentic AI systems are emerging in practice and research (Holldack et al. 2026; Allmendinger et al. 2026). With GenAI capabilities gradually entering the enterprise context via cloud-based enterprise platforms (with on-premise alternatives emerging) (Haki et al. 2025), and with agent modes from OpenAI and Google, multi-agent

frameworks such as Manus, and technologies like SAP's Joule or Microsoft's Agent Framework, agentic AI capabilities are becoming an integral part of enterprise platforms and can enable novel ways to integrate existing systems and technologies. Organizations can benefit from this development, as many previously unaddressed tasks span heterogeneous systems, creating a need for interoperability and requiring contextual interpretation that cannot be predefined at design time. Yet, realizing this potential requires first addressing challenges around governance, coordination, and human oversight.

Imagine a company that applies agentic AI systems at scale, e.g., a manufacturer of battery systems. It relies on a few critical suppliers for rare-earth materials. An agentic AI system continuously monitors environmental and geopolitical signals: a crawl agent scans news and regulations, detects floods in a key mining region, and triggers a supply risk-assessment goal. An agent for supplier relationship management (SRM) queries its enterprise resource planning (ERP) system to identify affected suppliers, open orders, and dependencies. Meanwhile, a customer relationship management (CRM) agent retrieves service-level agreements and critical customer commitments. A reporting agent compiles a structured assessment, quantifies revenue and delivery risk, and proposes mitigation strategies (e.g., alternative suppliers, SLA changes, repricing). The system prepares a briefing pack for the steering committee, and once human decision-makers approve a course of action, the agents autonomously initiate follow-up workflows in the organization's ERP systems, such as issuing requests for quotations (RFQs), updating risk scores, and drafting customer communications under predefined governance rules.

This example illustrates that agentic AI in enterprise settings often appears not as a single isolated artifact, but as

Accepted after two revisions by Oliver Krancher

M. M. Li (✉) · P. Reinhard · J. M. Leimeister
Chair for Information Systems, University of Kassel, Kassel,
Germany
e-mail: mahei.li@unisg.ch

P. Reinhard
e-mail: Philipp.reinhard@uni-kassel.de

J. M. Leimeister
e-mail: janmarco.leimeister@unisg.ch

M. M. Li · J. M. Leimeister
Institute of Information Systems and Digital Business (IWI-
HSG), University of St.Gallen, 9000 St. Gallen, Switzerland

a coordinated configuration of specialized agents embedded in heterogeneous systems. This raises the conceptual question of whether such systems should be understood as individual agents, multi-agent systems, or a distinct class of AI systems.

This ambiguity is also reflected in the literature. Despite growing interest and the emergence of various examples, definitions of agentic AI remain fragmented across practice and research. It remains unclear whether the term refers to individual AI agents, collections of interacting agents, or an entirely new class of systems relative to long-standing work on autonomous agents and multi-agent systems (MAS) (Wooldridge and Jennings 1995; Jennings and Wooldridge 1996). Building on this stream, recent IS work calls for research on multi-agent artificial intelligence (MAAI), proposing a multi-layered architecture (Allmendinger et al. 2026). Existing work on agentic AI varies, with some contributions treating agents merely as modular building blocks (Sapkota et al. 2025) and others reducing MAS to a purely architectural approach (Acharya et al. 2025). Additionally, Botti (2025) raises the question of whether current endeavors “reinvent the wheel”, since the foundational concept of agents was articulated decades ago. To address this ambiguity and connect with existing agent literature, this catchword article aims to provide a conceptualization of agentic AI and to delineate how contemporary agentic AI systems build upon and differ from earlier agent and MAS literature.

To this end, we first provide background on the technological evolution towards agentic AI, then define agentic AI and introduce first-order (individual AI agents) and second-order (multi-agent collaboration) agentic AI systems based on six core capabilities. We then discuss agentic workflows and conclude with a research agenda spanning design, governance, human-AI collaboration, and platform ecosystems.

2 Background

Agentic AI builds on, but is not defined by, recent advances in foundation models (Feuerriegel et al. 2024; Schneider et al. 2024). Early extensions to foundation models, such as retrieval-augmented generation (RAG), expanded model access to external information sources but remained limited to passive retrieval (Klesel and Wittmann 2025). In contrast, agentic AI leverages a set of complementary advances: tool learning (Qin et al. 2025; Qu et al. 2025), interaction protocols, such as the Model Context Protocol (MCP) and Agent-to-Agent (A2A) frameworks, which enable structured language-based coordination with systems and agents (Yang et al. 2025); and execution frameworks such as LAMs (Zhang et al. 2025), which enable

autonomous, context-aware actions across enterprise landscapes.

Rule-based systems, such as robotic process automation (RPA) (van der Aalst et al. 2018), excel at repetitive tasks but lack adaptability (Brynjolfsson and McAfee 2017). Classical machine learning (ML) and deep learning (DL) excel at pattern recognition but struggle with contextual goal alignment (Janiesch et al. 2021). GenAI and its underlying foundation models (Feuerriegel et al. 2024; Schneider et al. 2024) introduced flexible content generation without orchestration or action abilities. In contrast, agentic AI enables systems to autonomously interpret context, pursue goals, coordinate with other agents, and act through enterprise systems. Figure 1 illustrates this evolution as a layered structure, where each layer builds upon its predecessors rather than replacing them.

3 Defining Agentic AI

Building on recent AI developments (Feuerriegel et al. 2024; Schneider et al. 2024) and integrating streams from autonomous agents (Wooldridge and Jennings 1995), agentic information systems (IS) (Baird and Maruping 2021), and contemporary AI agent and agentic AI literature (Acharya et al. 2025; Shavit et al. 2023), we conceptualize agentic AI as a set of capabilities that enable AI-based information systems to autonomously perceive their environment, reason about their context, take goal-directed actions, and collaborate with other systems (Fig. 2). We refer to information systems that exhibit these capabilities as agentic AI systems, which we define as follows:

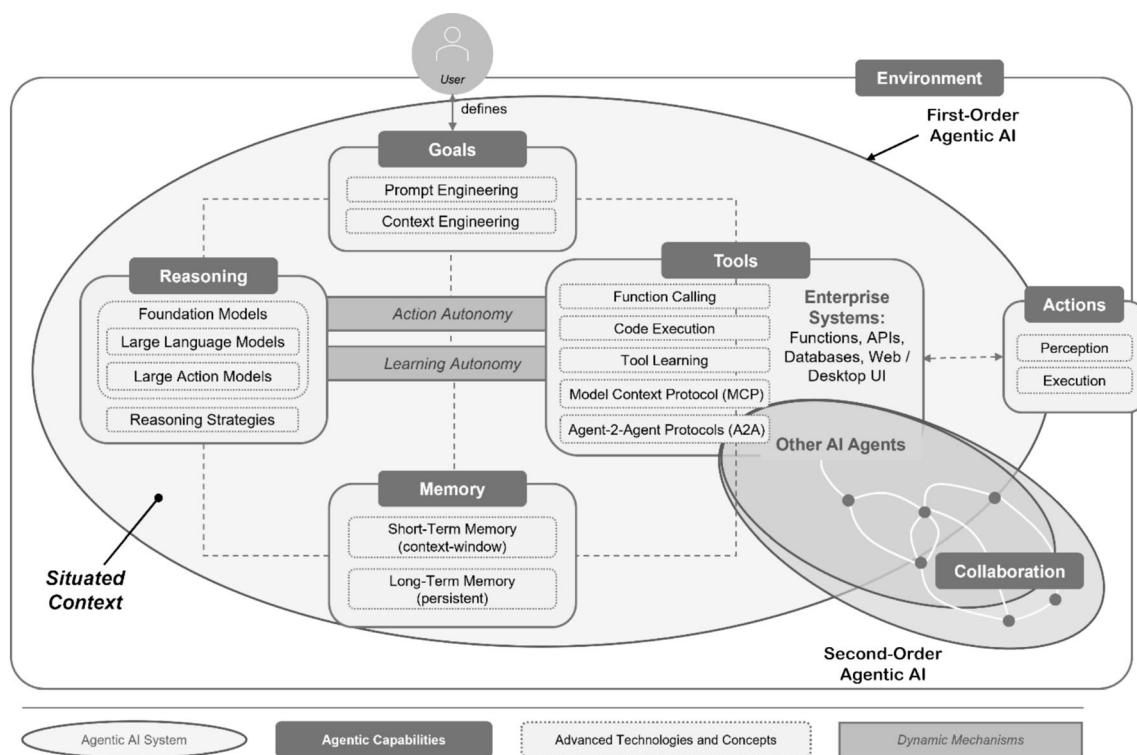
Definition: *Agentic AI systems consist of one (first-order) or multiple interacting AI agents (second-order) with the capabilities to observe their environment, learn to use and extend their tools, and autonomously reason when and how to act (i.e., through perception, planning, and execution) to achieve goals.*

Agentic AI systems respond to environmental changes, such as events or human input, by using (foundation-model-based) reasoning to determine whether action is required. Based on their memory and available tools, they identify suitable actions, evaluate which best fit the current context and goals, and execute actions accordingly. Second-order agentic AI systems exhibit the additional capability of collaboration among multiple first-order agentic AI systems.

A **first-order agentic AI** system is an AI-based system that exhibits six agentic capabilities (i.e., goals, environmental interaction, reasoning, memory, tools, and actions) and operates autonomously within its situated context. It perceives its environment, interprets these perceptions through reasoning processes utilizing foundation models

Fig. 1 Evolution of agentic AI

	Rule-Based Automation	Machine Learning and Deep Learning	Generative AI	Agentic AI
Primary Function and Use	Executing predefined workflows and scripts	Learning from data to predict, classify, or optimize outcomes	Generating novel content (text, image, code)	Acting autonomously to achieve goals in dynamic contexts
Technological Advancements	Symbolic reasoning through inference engines	Machine learning, deep learning, NLP, computer vision	Generative pre-trained transformer (GPT), foundation models, LLMs	Tool learning, large action and reasoning models, multi-agent collaboration
Knowledge Source for Learning	Hard-coded rules by experts	Predefined (ML) or learned (DL) features and criteria	Learned patterns from large-scale data	Dynamically inferred by interacting with enterprise systems
Human Involvement	Fully manual setup and rule definition	Data labeling, feature engineering, model training, and evaluation	Prompt engineering, fine-tuning, and human-in-the-loop validation	Prompt and context engineering, agent and workflow orchestration, oversight
Adaptivity	Operates only within predefined parameters	Generalizes from data but lacks real-time context awareness	Produces outputs conditioned on prompts and training context	Perceives, reasons, and acts based on evolving environmental states
Enterprise Integration	Limited and static workflows	Predefined enterprise processes and API-based integration	Ability to call functions and connect to enterprise data (RAG)	Autonomous discovery and coordination across enterprise tools, APIs, and agents
Exemplary Applications	Invoice processing, compliance checks	Spam filter, credit scoring, image recognition	Marketing content generation, coding	Sales agents, customer support agents
Available Solutions	UiPath, Blue Prism, SAP Business Rules Management	TensorFlow, PyTorch, AWS SageMaker	OpenAI (ChatGPT, DALL-E), Anthropic (Claude)	LangChain, AutoGPT, CrewAI, MCP

**Fig. 2** Conceptualizing agentic AI

(e.g., LLMs, LAMs), and takes actions based on them (Göldi and Rietsche 2024; Xi et al. 2025). Thus, by definition, a first-order agentic AI system is simply an *AI agent*. Throughout this paper, we use the term AI agent

synonymously with first-order agentic AI. Such AI agents can connect to other enterprise systems, learn and use tools, perform multiple tasks, and operate either alone or in collaboration with other AI agents.

A *second-order agentic AI* system or an agentic multi-agent system (*Agentic MAS*) is a configuration of two or more interacting AI agents in which *collaboration* transforms otherwise isolated AI agents (and potentially other systems or human actors) into a coherent system capable of shared or distributed goal pursuit. At least two agents must exhibit first-order agentic AI capabilities, enabling the system to perform tasks that span multiple agents' capabilities.

Our conceptualization of agentic AI systems builds on, yet differs from, three related literatures (See Online Appendix Table 2. The appendices are available via <http://link.springer.com>): (1) intelligent agents and MAS, (2) agentic IS, and (3) recent work on AI agents and agentic AI.

First, we draw on classical agent and MAS research (e.g., Wooldridge and Jennings 1995; Wooldridge 2009), which defines agents as autonomous software entities and MAS as coordinating societies of such entities. We retain this foundation by treating agentic AI as comprising both individual and multi-agent configurations, and extend it with a capability-based view that positions foundation models as the core enablers of adaptive reasoning, tool use, and system interaction. *Second*, we relate to the agentic IS literature (e.g., Russell 2019; Baird and Maruping 2021), which conceptualizes agentic IS as artifacts that autonomously act *on behalf of humans*. While we share this sociotechnical view of delegated autonomy and humans as goal definers, our conceptualization centers on the technological capabilities that enable agentic behavior in real enterprise settings. *Third*, we connect to recent work on AI agents and agentic AI (e.g., Acharya et al. 2025; Sapkota et al. 2025; Bandi et al. 2025), from which we synthesize emerging agentic capabilities. While this literature often emphasizes multi-agent constellations of agents (Sapkota et al. 2025), we argue that designing and studying agentic AI systems requires a differentiated understanding of agentic AI. We define both individual agents and MAS as agentic AI and distinguish first- and second-order forms operating across heterogeneous enterprise environments.

4 Six Core Agentic AI Capabilities

We synthesize insights from agent literature, agentic IS, and contemporary AI agent research to reveal six core agentic capabilities that characterize AI agents (first-order agentic AI), with Table 3 in the Online Appendix detailing their derivation across these streams.

4.1 Goal Capability

Goals are the objectives that AI agents pursue, with or without human oversight. In traditional AI systems and classical agents, such objectives are typically narrow and limited to well-defined tasks (e.g., summarization, translation, image recognition, or classification) that follow a simple input–output relation (Feuerriegel et al. 2024; Klesel and Wittmann 2025). In contrast, agentic AI systems can pursue one or multiple complex, long-term, and often open-ended goals, such as managing supply-chain risks, which require adaptive, contextual, and multi-step reasoning, tool use, and actions (Acharya et al. 2025).

Rather than being internal mental states, goals are formed and pursued within specific socio-material contexts (Baird and Maruping 2021). They remain aligned with human intentions and are operationalized through reasoning and execution based on the situated context. Beyond explicitly delegated objectives, agentic AI systems can also derive operational goals from events or human prompts, interpreting these inputs as environmental triggers that initiate reasoning. In LLM-based agents, goals are typically expressed in natural language and instantiated through *prompt* and *context engineering* or hybrid natural-language/JSON formats that planner agents decompose into actionable subgoals and subtasks.

Example: In the supply-chain scenario, the high-level goal of ensuring resilience is decomposed into subgoals such as assessing supplier exposure, evaluating customer impact, and identifying mitigation strategies. The AI agent may pursue these subtasks autonomously, utilize available tools, or delegate work to other agents.

4.2 Environment Interaction Capability

The environment refers to the dynamic, often complex external conditions in which agentic AI systems operate and perform actions (Wooldridge and Jennings 1995). It may be physical, digital, or socio-technical; and encompasses other agents, humans, enterprise systems, and event streams such as regulatory updates or workflow changes (Acharya et al. 2025; Shavit et al. 2023; Wang et al. 2023). Environmental interaction describes how agents perceive and respond to these conditions. While the environment defines the overall space in which an AI agent exists, only a subset of this space is directly perceivable and internally representable (Wooldridge 2009). This *situated context* represents the portion of the environment that the AI agent can perceive, reason about, and act upon. It is shaped by available tools, system interfaces, permissions, and

integration points that allow agents to observe signals or execute actions. Thus, tools determine not only how an AI agent acts but also which parts of the environment become perceivable and executable.

Environments both constrain and enable agentic behavior. They provide informational input, define available actions (Baird and Maruping 2021), and impose boundary conditions, such as policies, contracts, governance rules, and system states, that shape how agents reason and act. This parallels how social, organizational, and material contexts shape human behavior. Enterprise systems (e.g., ERP, CRM, SRM, RPA) are part of the external environment in which the AI agent operates. When exposed via APIs or functions, however, they become tools that extend the agent's situated context and its actions. Advances in foundation models expand the situated context by enabling richer perception (e.g., multimodal understanding) and reasoning, and by supporting tool-mediated access to heterogeneous enterprise systems.

Example: The environment of the above-discussed agentic AI system includes real-time news, weather, and regulatory information; enterprise systems such as CRM or SRM; and human actors involved in procurement and sales. The agent's situated context is the portion it can access through tools, for example, flooding alerts, relevant supplier and customer data, available procurement APIs, and interactions with other agents.

4.3 Tool Capability

Tools are external modules, such as APIs, applications, databases, software functions, or other AI agents, that enable AI agents to perceive their environment, extend their reasoning, and execute actions (Wang et al. 2023; Qin et al. 2025). Tools form the operational interface between the AI agent and its environment: they determine which data the AI agent can access, which system states it can read or modify, and which actions it can execute.

Access to tools, therefore, defines the agent's action space, both in terms of perception and execution. As agents gain access to additional tools, they can pursue increasingly complex goals, interact with a broader range of systems, and act with greater autonomy. In enterprise settings, tools typically include search and retrieval APIs, database query interfaces for CRM or ERP data, email or calendar APIs, workflow functions, RPA bots, and other AI agents that can be invoked as modular components.

Agentic AI systems increasingly employ techniques such as function calling, code execution, or dynamic tool creation (e.g., writing and executing Python code) to extend their toolsets (Wölflein et al. 2025; Göldi and

Rietsche 2024). Beyond *tool creation*, modern agentic AI systems also learn how to use tools, a capability often referred to as *tool learning* (Qin et al. 2025). Through iterative reasoning, agents learn to select appropriate tools, parameterize (e.g., SQL queries, search prompts, API payloads), chain multiple tools into multi-step workflows, and adjust their plans based on intermediate results (Qin et al. 2025).

Emerging protocols such as the MCP and A2A frameworks further enhance this capability by standardizing tool discovery, access, and interaction. MCP exposes heterogeneous enterprise tools through unified interfaces, while A2A protocols allow AI agents to treat other AI agents as callable tools, coordinating agentic MAS.

Example: In the supply-chain scenario, tools include: ERP, SRM, and CRM APIs for retrieving supplier, order, and contract data, web-crawling and search APIs for detecting disruptions, and workflow or RPA functions for issuing RFQs, updating risk scores, or generating customer communications.

4.4 Reasoning Capability

Reasoning is the core capability that enables agentic AI systems to decompose high-level objectives into multi-step operational plans and actions in dynamic environments (Shavit et al. 2023; Acharya et al. 2025). Agentic AI systems employ iterative, context-sensitive reasoning processes to decide when and how to act – decomposing goals, selecting appropriate tools, and sequencing actions based on feedback (e.g., Lee et al. 2024; Qin et al. 2025). These processes are supported by reasoning strategies, for example Chain-of-Thought (Wei et al. 2022), ReAct (Yao et al. 2023a), Tree-of-Thought (Yao et al. 2023b), or Reflexion (Shinn et al. 2023), which structure how agents decompose goals, select and orchestrate tools, draw on memory, and determine their action space.

Contemporary implementations commonly rely on foundation models and derivatives (e.g., LLMs, LAMs) to support this reasoning capability (Feuerriegel et al. 2024). Here, the concept of reasoning itself, however, remains model-agnostic. While current model-based agents may exhibit reasoning behaviors that resemble aspects of human deliberation, they do so through inference (patterns) rather than replicating genuine human cognition. Advances in large-scale AI models enhance practical capabilities in goal decomposition, task prioritization, and real-time decision-making across enterprise systems (Feuerriegel et al. 2024).

Example: In the supply-chain scenario, the LLM-based AI agent interprets the disruption data, infers its likely impact, decomposes the goal into steps such as querying

ERP or checking CRM commitments, and adjusts its plan iteratively as new information becomes available.

4.5 Action Capability

Actions in agentic AI refer to autonomous, goal-directed operations that affect the environment (cf. Acharya et al. 2025; Shavit et al. 2023; Qin et al. 2025). We distinguish between two main types of actions: *perceptions*, such as collecting data from sensors, and *executions*, such as updating records in CRM or ERP systems or generating and dispatching reports or emails. These actions result from the AI agent's use and orchestration of tools. As such, they include the activation of other agents. Tools define the set of operations an AI agent can perform, but which actions become relevant or are selected are based on the agent's goals and contextual interpretation and thus determined by its reasoning processes.

The level of autonomy varies depending on the degree to which the AI agent can autonomously prioritize, initiate, and execute actions, which range from decision support with a human-in-the-loop or human-on-the-loop, to fully automated execution (Baird and Maruping 2021; Murray et al. 2021), to self-evolving AI agents (Fang et al. 2025). Recent developments focus on LAMs that equip agentic AI systems with specialized operational capabilities, such as web automation or robotic control, enhancing both precision and task autonomy (Zhang et al. 2025). LAMs more tightly integrate reasoning, tools, and actions within the underlying foundation models.

Example: In the supply-chain scenario, a procurement agent turns its plan into concrete actions by issuing RFQs via an RPA function, updating supplier risk scores in SAP, and generating customer notifications for human approval. These actions are realized through composed tool calls, allowing the AI agent to change system states and trigger workflows with a high degree of operational autonomy.

4.6 Memory Capability

Memory provides the internal structures through which agentic AI systems build and maintain their *situated context*, the actionable subset of the environment the AI agent can perceive, reason about, and use to execute actions. Memory enables continuity across interactions, allowing the AI agent to relate new observations to prior states, refine its understanding, and support multi-step reasoning.

Agentic AI systems rely on memory structures to maintain goals, intermediate results, tool outputs, and contextual information over time. Short-term memory (e.g., context windows, working buffers) provides continuity within an interaction, while long-term memory (e.g., vector stores, knowledge graphs) preserves relevant

information across episodes and tasks. Additionally, reinforcement learning from (human) feedback enables models to internalize experiences through iterative fine-tuning, effectively reshaping their behavior over time. Together, these mechanisms enable autonomous adaptation and long-horizon reasoning, extending the agent's capacity beyond what isolated inputs could support (Zhang et al. 2024; Wang et al. 2023).

Memory serves as more than a repository of past information. It supports reasoning by helping the AI agent interpret its goals, assess its context, and choose appropriate actions. By linking new observations to prior states, memory provides the contextual basis for interpreting environmental inputs.

Example: In the supply-chain scenario, a risk-assessment agent stores past disruptions, supplier vulnerabilities, and mitigation strategies using a vector database. When a new flood event occurs, the AI agent retrieves similar past cases to contextualize likely impacts, refine risk scores, and propose appropriate procurement actions, enabling consistent reasoning beyond a single AI response.

4.7 Action and Learning Autonomy

Beyond these structural capabilities, agentic AI systems rely on two dynamic mechanisms: *learning autonomy* and *action autonomy*. These mechanisms define how agents adapt to their environment and act within it by updating their situated context and how independently they can prioritize and execute actions (Hosseini and Seilani 2025).

Learning autonomy is an agent's ability to update its contextual understanding in response to environmental changes. This spans a continuum from static learning (e.g., pre-training) to dynamic adaptation (e.g. real-time inputs) (Queiroz et al. 2024). Central to learning autonomy is the agent's capacity to update its memory and incorporate new tools, thereby refining its reasoning processes and expanding its future action repertoire. In the supply-chain scenario, learning autonomy enables the risk-assessment agent to refine supplier risk scores and mitigation strategies over time as it observes repeated or evolving disruption patterns.

Action autonomy refers to the extent to which an AI agent can independently select and execute actions. This ranges from augmentation, where humans make final decisions, to full automation, where the AI agent acts with minimal oversight (Murray et al. 2021). Action autonomy depends on the agent's ability to perceive context, interpret goals, and translate plans into executable workflows. In the supply-chain example, action autonomy means the AI agent can issue RFQs, adjust supplier risk ratings in SAP, or draft customer communications once a disruption is

detected, requiring human approval only in predefined cases.

As AI agents act, their reasoning processes may identify additional tool options or refine how available tools can be combined, creating a continuous feedback loop between learning and execution. Together, learning and action autonomy enable agentic systems to sustain adaptive behavior under evolving conditions by aligning and extending perception, memory, reasoning, and tool-mediated action (Queiroz et al. 2024), moving the field of agentic AI towards self-evolving systems (Fang et al. 2025).

An AI agent equipped with these capabilities can autonomously pursue goals within its situated context. However, enterprise environments typically involve distributed responsibilities, heterogeneous systems, and interdependent tasks. In such settings, agents do not operate in isolation. Instead, complex goals are pursued through the interaction of multiple agents, which form what we conceptualize as second-order agentic AI and serve as the focus of the following section.

5 Multi-agent Collaboration

Second-order agentic AI refers to configurations of multiple agents whose collaboration creates capabilities that no individual AI agent possesses on its own. These capabilities arise when AI agents share or recombine goals, tools, memory, and actions, allowing the overall system to solve composite tasks, cross-validate information, and coordinate actions toward a common goal with distributed sub-goals in ways that exceed any individual agent's action space. Early MAS were typically designed as expert systems with predefined roles, static coordination logic, and limited autonomy to adapt to unforeseen conditions (McArthur et al. 2007; Wooldridge 2009). While their architectures and protocols remain relevant, agentic MAS extend these capabilities, enabled by foundation-model-based reasoning and tool-use, to negotiate, delegate, and replan through iterative reasoning processes communicated via natural language.

Such agentic MAS build on these advances by treating **collaboration** not as a fixed architecture (Acharya et al. 2025) but as an emergent capability that arises from the interaction of multiple AI agents. This view resonates with the idea that collective configurations can give rise to new behavioral capacities (Minsky 1986) and aligns with recent work on generative agent societies (Park et al. 2023). Two conceptual design challenges are central to second-order agentic AI: First, second-order agentic AI systems exhibit a recursive structure in which agentic capabilities (goals, environmental interaction, reasoning, memory, tools,

actions) are instantiated both at the level of individual agents and within the collective configuration. Second, collaboration depends on coordination structures that shape how agents share goals, distribute sub-tasks, and align actions. Second-order agentic AI, therefore, requires additional mechanisms that govern how agents collaborate, align, and exchange information.

The current field of multi-agent archetypes and orchestration is heterogeneous and rapidly evolving. To clarify how these approaches differ from conventional MAS paradigms (Sujil et al. 2018; McArthur et al. 2007), we present two illustrative archetypes:

- (1) **Centralized:** Centralized agentic AI systems follow a hub-and-spoke configuration, where a central AI agent handles planning and coordination. It decomposes goals into subtasks and assigns them to peripheral agents, which execute specific functions (Sujil et al. 2018). Currently, it is the most dominant approach to multi-agent orchestration. A structured subtype of this architecture is the hierarchical form, in which the central planner does not delegate to a single flat layer of agents but to multiple layers that mirror organizational or process hierarchies. In such designs, higher-level agents translate strategic goals, mid-level agents create operational subplans, and lower-level agents execute actions within enterprise systems (Shah et al. 2024; Izmirliglu et al. 2024). Unlike classical MAS with rigid task routing, modern coordinators use LLM-based reasoning to dynamically decompose tasks, reassign responsibilities, and adjust plans based on intermediate results. In centralized second-order systems, system-level autonomy emerges from an individual coordinating AI agent that integrates and aligns the actions of other agents. Example: *A central orchestrator agent receives the flood alert, breaks the task into steps, and assigns them to crawler, SRM, CRM, and reporting agents and then coordinates the final updates, such as issuing RFQs or adjusting risk scores.*
- (2) **Decentralized:** In decentralized configurations, collaboration is achieved through distributed negotiation and shared decision-making among agents. Agents coordinate peer-to-peer, performing reasoning and actions based on local observations and shared information (Sujil et al. 2018; Shah et al. 2024). LLM-enabled agents enhance this pattern by enabling agents to negotiate goals, interpret intent, and resolve conflicts via natural language (Liu et al. 2025). This structure supports scalability and robustness but requires mechanisms to prevent inconsistencies and handle nondeterminism. Example: *After*

a crawler agent shares the alert, SRM, CRM, logistics, and pricing agents each analyze the impact using their own data and exchange natural-language messages to agree on mitigation actions without a central controller.

6 Agentic Workflows

Agentic AI is increasingly entering enterprise platforms (e.g., Microsoft Azure, Google Cloud), which now offer agent SDKs and low-code workflow orchestrators (e.g., LangChain, n8n, Azure AI Foundry) for integrating AI agents. Recognizing initial work on agentic workflows (Sapkota et al. 2025; Allmendinger et al. 2026), we draw on our capability-based conceptualization to demonstrate how agentic AI can be integrated into workflow settings. Based on our definition, a workflow qualifies as agentic when it incorporates at least one first-order agent. Workflows that only call a LLM-model for a single predefined task do not meet these conditions. Figure 3 shows two types of workflows that integrate AI agents and compares them to traditional and non-agentic workflows.

- (1) *Automated workflows (non-AI)* are deterministic and rule-driven. They execute predefined steps and scripts (van der Aalst et al. 2018) without any AI component that represents goals, performs multi-step reasoning, selects tools, or relies on memory (e.g., triggering a fixed rule when a shipment is delayed).

- (2) *AI workflows (non-agentic)* call an AI system as a tool inside a single step (e.g., using an LLM once to rewrite a disruption message or a RAG system for document retrieval) (Janiesch et al. 2021; Feuerriegel et al. 2024; Klesel and Wittmann 2025), but without agentic capabilities.
- (3) *First-order agentic workflows* contain an individual AI agent that integrates goals, iterative reasoning, tool use, memory, and action execution to manage complex, multi-step tasks (e.g., an AI agent that receives a flood alert, plans supplier checks, and triggers follow-up actions). The AI agent decides which actions to take to achieve the process goals rather than performing specific and predefined tasks (Vu et al. 2025).
- (4) *Second-order agentic workflows* involve multiple AI agents working in the same orchestration environment (Sapkota et al. 2025; Allmendinger et al. 2026). Each AI agent performs its own reasoning and tool use, and the agents coordinate task boundaries, exchange relevant information, and synchronize memory where needed (e.g., a risk agent flagging exposed suppliers while a procurement agent evaluates alternatives). This enables more complex, adaptive, and goal-driven workflows.

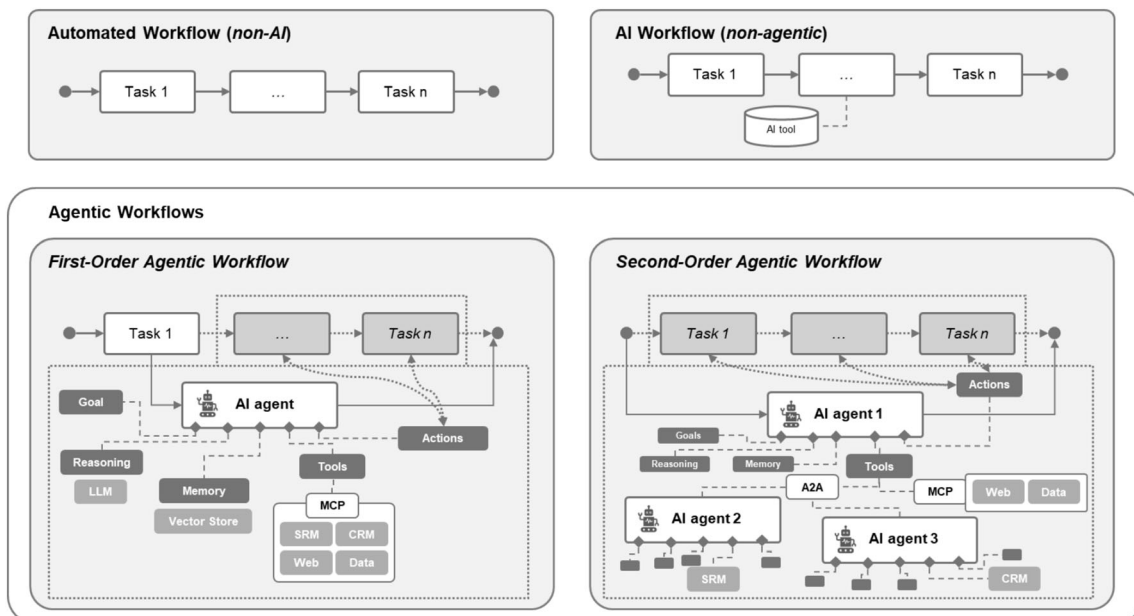


Fig. 3 Agentic AI workflows

7 Discussion of Areas for Future IS Research

The emergence of agentic AI expands the scope of AI systems in enterprises (Maedche et al. 2019; Feuerriegel et al. 2024; Janiesch et al. 2021) by introducing autonomous, tool-using AI agents and multi-agent constellations capable of pursuing complex goals across system landscapes. Building on the conceptual foundations, architectures, and challenges outlined above, these developments open an agenda for IS research.

Table 1 presents exemplary research questions across five key directions, spanning technological, organizational, and societal implications of agentic AI.

7.1 Designing Agentic AI Systems

A first research avenue concerns the design of agentic AI systems, focusing on the technical and architectural choices that configure AI agents and extends them into multi-agent constellations. This includes the design of the operational infrastructure (e.g. an agent harness) that instantiates an AI agent's core capabilities, such as tool execution, memory management, and reasoning orchestration, as well as approaches that leverage AI agents to support the system design process itself (e.g. agentic engineering). At the level of individual agents, research is needed on mechanisms that enable agents to detect ambiguous or underspecified

Table 1 Exemplary research questions for future IS research on agentic AI

Research direction	Exemplary research questions	Related IS literature
Designing Agentic AI Systems	How can agentic AI systems be designed to handle ambiguous goals, request clarification, and structure complex tasks effectively? What (model) representations best support memory for enterprise agents? How can agents learn new tools autonomously? How should agentic AI systems be designed to coordinate and cooperate reliably? How can the performance of different agentic AI configurations be measured and compared?	e.g., He et al. (2025), Klesel and Wittmann (2025), Göldi and Rietsche (2024), Allmendinger et al. (2026)
Agentic Workflows and Process Management (Agentic BPM)	How can agentic workflows be modelled to represent autonomous goal decomposition and task coordination? How can reliable execution of agentic workflows be ensured under dynamic conditions? How should agentic workflows be governed when multiple agents interact with enterprise systems and organizational actors?	e.g., Fettke and Di Francescomarino (2025), Rosemann et al. (2024), Vu et al. (2025)
Towards Agentic AI Governance	How can emergent affordances of agentic AI systems be governed? How should governance structures adapt when decision rights intertwine with second-order agentic AI architectures? Which mechanisms ensure oversight and accountability as agentic AI systems gain decision authority?	e.g., Berente et al. (2021), Birkstedt et al. (2023), Shavit et al. (2023), Saunders et al. (2020)
Human-Agentic AI Collaboration and Delegation	When do humans delegate or reclaim decision rights from agentic AI systems? How do feedback and experience shape human reliance on agentic AI in hybrid decision settings? Which forms of explanation best support human oversight and appropriate overrides in second-order agentic AI systems?	e.g., Maedche et al. (2019), Strunk et al. (2024), Baird and Maruping (2021), Fügener et al. (2022), Adam et al. (2024)
Agentic Web and Agentic AI Platform Ecosystems	When and why do platform providers choose to enclose or open agentic orchestration layers? What governance structures enable cross-platform value creation in fragmented agentic ecosystems? How are platform business models transformed when agentic AI systems act on behalf of users across digital platforms?	e.g., Hein et al. (2020), Wessel et al. (2025), Haki et al. (2025), Heimbürg et al. (2025)

goals, request clarification, and decompose tasks reliably. A central challenge is designing short-term, long-term, and shared memory architectures that balance adaptability with stability and provide an auditable context for reasoning. This includes identifying which internal representations (e.g., vector stores, knowledge graphs, episodic logs) best support transparent and verifiable agent behavior (Klesel and Wittmann 2025; Wang et al. 2023). Tool-use design also remains critical, as agents must learn new tools and action affordances while staying aligned with system boundaries and intended capabilities (Göldi and Rietsche 2024; Qin et al. 2025).

Beyond individual agents, designing second-order agentic AI requires understanding when single-agent configurations suffice and when multi-agent setups yield advantages under varying levels of task complexity and uncertainty (Allmendinger et al. 2026). Research should examine which architectural archetypes influence how agentic MAS dynamically reconfigure as agents join, leave, or shift roles to achieve their shared goals. Because LLM-based agents increasingly coordinate through natural-language interaction, a further design challenge concerns how to structure and constrain inter-agent communication and negotiation to prevent miscommunication, drift, or collusion (Guo et al. 2024; Hu et al. 2025).

7.2 Agentic Workflows and Business Process Management

Agentic workflows extend traditional business process management (BPM) by enabling autonomous goal decomposition, adaptive task coordination, and context-sensitive execution across enterprise processes. This shifts BPM from predefined control-flow automation toward more flexible, goal-driven, and autonomous forms of orchestration, such as agentic workflows (Rosemann et al. 2024; Grisold et al. 2024) as shown in Fig. 3. Enterprise systems begin to expose such capabilities through emerging agentic interfaces and orchestration tools (e.g., Microsoft Agent Framework; Salesforce Agentforce).

These developments give rise to *agentic BPM* (Vu et al. 2025) which uses autonomous AI agents as process actors and agent-based abstractions to analyze how they plan, coordinate, and execute activities. This raises new challenges for how agentic workflows should be modelled, verified, and governed (Amirkhani and Barshooi 2022; Shi et al. 2024), particularly when agents decompose goals into sequenced or parallel subtasks (He et al. 2025; Yue et al. 2024) or interact with other agents and organizational actors.

7.3 Towards Agentic AI Governance

AI systems challenge existing models of IS and AI governance (Berente et al. 2021; Gregory et al. 2018; Birkstedt et al. 2023). This holds for agentic AI. In prior IS governance models (Saunders et al. 2020), IT-related decision rights were distributed and delegated among human actors (Gregory et al. 2018). In contrast, emerging AI agents can autonomously generate code and interact with multiple enterprise systems (Wölflein et al. 2025; Göldi and Rietsche 2024). As a result, the scope of governance expands beyond IT departments, end-users, and other stakeholders to include agentic AI systems themselves. Thus, governance needs to address emergent affordances: agents may develop capabilities not foreseen at implementation (Leonardi 2011; Strong et al. 2014; Beck et al. 2022). This is particularly relevant for LLM-based systems, where reasoning and consensus processes are often opaque and require runtime supervision (He et al. 2025; Shi et al. 2024).

Likewise, structural governance mechanisms, such as the degree of centralization or decentralization of IT decision rights, become tightly intertwined with the architectural design of second-order agentic AI systems (He et al. 2025; Allmendinger et al. 2026), where agents not only perform tasks but also influence how IT capabilities evolve. As agents learn, delegate, and act across organizational boundaries, new challenges arise around oversight, control, and alignment with regulatory and enterprise objectives. As agentic AI systems gain decision authority, responsibilities, and control shift across human–AI configurations (Söllner et al. 2025; Gregory et al. 2018).

Researchers and practitioners must also contend with risks exacerbated by the non-deterministic nature of agentic workflows and their tight integration into enterprise systems. Agentic AI inherits the stochastic behavior, hallucinations, and bias propagation of LLMs (Guo et al. 2024) and further amplifies them through noisy perceptions, tool-mediated actions, and emergent multi-agent phenomena such as collusion or knowledge drift (Hu et al. 2025). This layered uncertainty undermines reproducibility, complicates validation, and limits the usefulness of existing uncertainty quantification methods (Abbasli et al. 2025). Future governance models must delineate clear responsibilities for oversight and accountability, ensuring regulatory compliance, mitigating risks, and embedding accountability into both design-time structures and runtime supervision. This includes addressing who is accountable for the agentic AI systems that exercise governance functions.

7.4 Human-Agentic AI Collaboration and Delegation

This research avenue explores how control, responsibility, and task ownership are distributed across humans and agentic AI systems in non-dyadic constellations (Baird and Maruping 2021; Reinhard et al. 2026). Agentic AI enables decision rights to shift dynamically, between agents, from agents to humans, or across roles, based on context, capability, and coordination logic. Understanding these dynamics is essential for determining under what conditions humans delegate, share, or reclaim decision rights (Fügener et al. 2022; Strunk et al. 2024) from agentic AI systems in enterprise processes.

Agentic AI-based collaboration relies on distributed decision mechanisms, such as negotiation, voting, and argumentation among agents (Amirkhani and Barshooi 2022; Liu et al. 2025; Wooldridge 2009), which are increasingly intertwined with human decision-making. These mechanisms reflect the core logic of second-order agentic AI. Such non-dyadic decision settings raise essential questions about how humans develop appropriate reliance on agentic AI recommendations, and how feedback, experience, and learning over time recalibrate that reliance.

However, LLM-based reasoning often remains opaque, especially in consensus formation. This underscores the need for explainability and observability (Kim et al. 2025; Steyvers et al. 2025). Consequently, new research avenues emerge to investigate what forms of explanation, such as chain-of-thought, confidence scores, or conflicting agent states, best support human trust calibration and enable appropriate overrides of agentic AI decisions. Expanding explainable GenAI (GenXAI) to explainable agentic AI (*agentic XAI*) will be critical for ensuring traceability and reliability in multi-agent collaboration. Finally, assigning humans or agents to roles such as advisor, gatekeeper, or autonomous actor reshapes organizational perceptions of control and responsibility (Baird and Maruping 2021; Strunk et al. 2024; Adam et al. 2024).

7.5 Agentic AI in Platform Ecosystems

Platform research shows that value creation concentrates at points where complementors are coordinated and interactions are governed (Hein et al. 2020; Tiwana 2013; Park et al. 2023). Recent work demonstrates that GenAI shifts platform control upward toward interaction and orchestration layers (Wessel et al. 2025). Agentic AI is expected to amplify this trend as platform providers attempt to position agents, instead of APIs, as user-facing interfaces or to anchor enterprise activity within their own orchestration environments (e.g., agentic workflow platforms). This extends earlier work on platform enclosure and boundary

control (Ghazawneh and Henfridsson 2013; Eaton et al. 2015), and raises questions about openness, interoperability, and access rights in environments where autonomous agents interact across organizational and platform boundaries. As agent-ready enterprise landscapes emerge, current vendor strategies risk fragmenting the agentic ecosystem and limiting cross-platform value creation (Heimburg et al. 2025).

Beyond enterprise settings, the idea of Internet of Agents (IoA) (Yu et al. 2024) introduces environment where autonomous agents are benchmarked on complex web interactions. These experiments show that agents can already perform tasks resembling those of customers or intermediaries, raising new questions about platform roles, pricing structures, and access rights in agent-mediated digital ecosystems (Heimburg et al. 2025; Eaton et al. 2015).

8 Conclusion

Agentic AI extends contemporary AI systems through six core capabilities – goals, environment interaction, tools, memory, reasoning, and actions – that enable AI-based information systems to autonomously perceive, plan, and act within their situated context. Building on this capability-based view, we distinguished first-order agentic AI (individual AI agents) from second-order agentic AI (multi-agent collaboration). We further illustrate how this capability-based definition can differentiate traditional workflow automation types from agentic (AI) workflows.

The interplay of action and learning autonomy, orchestrated primarily through reasoning capability, enables agentic AI systems not only to adapt but to continuously self-improve. At the frontier, autonomous research agents (Karpathy 2026) exemplify this trajectory: agentic AI systems that iteratively experiment, evaluate outcomes, and optimize themselves.

Finally, we build on this conceptual foundation to outline a research agenda spanning the design of agentic AI systems, agentic workflows and BPM, agentic AI governance, human–agentic AI collaboration and delegation, and platform ecosystems. Taken together, these directions provide a structured basis for future research on agentic AI systems within a socio-technical enterprise context.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12599-026-01009-w>.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- van der Aalst WMP, Bichler M, Heinzl A (2018) Robotic process automation. *Business & Information Systems Engineering* 60(4):269–272
- Abbasli T, Toyoda K, Wang Y, Witt L, Ali MA, Miao Y, et al (2025) Comparing uncertainty measurement and mitigation methods for large language models: a systematic review. In [arXiv:2504.18346](https://arxiv.org/abs/2504.18346)
- Acharya DB, Kuppan K, Divya B (2025) Agentic AI: autonomous intelligence for complex goals — a comprehensive survey. *IEEE Access* 13:18912–18936
- Adam M, Diebel C, Goutier M, Benlian A (2024) Navigating autonomy and control in human-AI delegation: user responses to technology-versus user-invoked task allocation. *Decis Support Syst* 180:114193
- Allmendinger S, Bonenberger L, Endres K, Fetzner D, Gimpel H, Kühl N (2026) Multi-agent AI. *Electron Mark*. <https://doi.org/10.1007/s12525-025-00862-z>
- Amirkhani A, Barshooi AH (2022) Consensus in multi-agent systems: a review. *Artif Intell Rev* 55(5):3897–3935
- Baird A, Maruping L (2021) The next generation of research on IS use: a theoretical framework of delegation to and from agentic IS artifacts. *MIS Q* 45(1):315–341
- Bandi A, Kongari B, Naguru R, Pasnoor S, Vilipala SV (2025) The rise of agentic AI: a review of definitions, frameworks, architectures, applications, evaluation metrics, and challenges. *Future Internet* 17(9):404
- Beck R, Dibbern J, Wiener M (2022) A multi-perspective framework for research on (sustainable) autonomous systems. *Bus Inf Syst Eng* 64(3):265–273
- Berente N, Gu B, Recker J, Santhanam R (2021) Managing artificial intelligence. *MIS Q*. <https://doi.org/10.25300/misq/2021/16274>
- Birkstedt T, Minkinen M, Tandon A, Mäntymäki M (2023) AI governance: themes, knowledge gaps and future agendas. *Internet Res* 33(7):133–167
- Botti V (2025) Agentic AI and multiagentic: are we reinventing the wheel? In [arXiv:2506.01463](https://arxiv.org/abs/2506.01463)
- Brynjolfsson E, McAfee A (2017) Artificial intelligence, for real. *Harv Bus Rev* 1(1):1–31
- Eaton B, Elaluf-Calderwood S, Sørensen C, Yoo Y (2015) Distributed tuning of boundary resources. *MIS Q* 39(1):217–244
- Fang J, Peng Y, Zhang X, Wang Y, Yi X, Zhang G, et al (2025) A comprehensive survey of self-evolving ai agents: a new paradigm bridging foundation models and lifelong agentic systems. In [arXiv:2508.07407](https://arxiv.org/abs/2508.07407)
- Fettke P, Di Francescomarino C (2025) Business process management and artificial intelligence. *KI-Künstl Intell* 39:1–13
- Feuerriegel S, Hartmann J, Janiesch C, Zschech P (2024) Generative AI. *Bus Inf Syst Eng* 66(1):111–126. <https://doi.org/10.1007/s12599-023-00834-7>
- Fügner A, Grahl J, Gupta A, Ketter W (2022) Cognitive challenges in human-artificial intelligence collaboration: investigating the path toward productive delegation. *Inf Syst Res* 33(2):678–696
- Ghazawneh A, Henfridsson O (2013) Balancing platform control and external contribution in third-party development: the boundary resources model. *Inf Syst J* 23(2):173–192
- Göldi A, Rietsche R (2024) Making sense of large language model-based AI agents. In : *ICIS 2024*
- Haki K, Safaei D, Magan A, Griffiths M (2025) Integrating generative AI into enterprise platforms: insights from Salesforce. *Inf Syst J*. <https://doi.org/10.1111/isj.12593>
- Hu J, Dong Y, Ao S, Li Z, Wang B, Singh L, et al (2025) Position: towards a responsible LLM-empowered multi-agent systems. In [arXiv:2502.01714](https://arxiv.org/abs/2502.01714)
- Gregory RW, Kaganer E, Henfridsson O, Ruch TJ (2018) IT consumerization and the transformation of IT governance. *MIS Q* 42(4):1225–1229
- Grisold T, Kremser W, Mendling J, Recker J, vom Brocke J, Wurm B (2024) Generating impactful situated explanations through digital trace data. *J Inf Technol* 39(1):2–18
- Guo T, Chen X, Wang Y, Chang R, Pei S, Chawla NV, et al (2024) Large language model based multi-agents: a survey of progress and challenges. In: *Proceedings of the thirty-third international joint conference on artificial intelligence*, pp 8048–8057
- He J, Treude C, Lo D (2025) LLM-based multi-agent systems for software engineering: literature review, vision and the road ahead. *ACM Trans Softw Eng Methodol* 34(5):1–30
- Heimbürg V, Schreieck M, Wiese M (2025) Complementor value co-creation in generative AI platform ecosystems. *J Manag Inf Syst* 42(2):491–528
- Hein A, Schreieck M, Riasanow T, Setzke DS, Wiese M, Böhm M, Krcmar H (2020) Digital platform ecosystems. *Electron Mark* 30:87–98
- Holldack F, Banh L, Strobel G (2026) Agentic information systems. *Electron Mark*. <https://doi.org/10.1007/s12525-025-00861-0>
- Hosseini S, Seilani H (2025) The role of agentic AI in shaping a smart future: a systematic review. *Array (Amst)*. <https://doi.org/10.1016/j.array.2025.100399>
- Izmirlioglu Y, Pham L, Son TC, Pontelli E (2024) A survey of multi-agent systems for smartgrids. *Energies* 17(15):3620. <https://doi.org/10.3390/en17153620>
- Janiesch C, Zschech P, Heinrich K (2021) Machine learning and deep learning. *Electron Mark* 31(3):685–695
- Jennings N, Wooldridge M (1996) Software agents. *IEE Rev* 42(1):17–20
- Karpathy, Andrej (2026) Autoresearch: AI agents running research on single-GPU nanochat training automatically. GitHub
- Kim SSY, Vaughan JW, Liao QV, Lombrozo T, Russakovsky O (2025) Fostering appropriate reliance on large language models: the role of explanations, sources, and inconsistencies. In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*
- Klesel M, Wittmann HF (2025) Retrieval-augmented generation (RAG). *Bus Inf Syst Eng* 67:551–561. <https://doi.org/10.1007/s12599-025-00945-3>
- Lee Y, Ferber D, Rood JE, Regev A, Kather JN (2024) How AI agents will change cancer research and oncology. *Nat Cancer* 5(12):1765–1767. <https://doi.org/10.1038/s43018-024-00861-7>
- Leonardi PM (2011) When flexible routines meet flexible technologies: affordance, constraint, and the imbrication of human and material agencies. *MIS Q* 35(1):147–167
- Liu F, Dai W, Zhang C, Zhu J, Yao L, Li X (2025) Co-LLaVA: efficient remote sensing visual question answering via model

- collaboration. *Remote Sens* 17(3):466. <https://doi.org/10.3390/rs17030466>
- Maedche A, Legner C, Benlian A, Berger B, Gimpel H, Hess T et al (2019) AI-based digital assistants. *Bus Inf Syst Eng* 61(4):535–544
- McArthur SDJ, Davidson EM, Catterson VM, Dimeas AL, Hatziar-gyriou ND, Ponci F, Funabashi T (2007) Multi-agent systems for power engineering applications — part I: concepts, approaches, and technical challenges. *IEEE Trans Power Syst* 22(4):1743–1752. <https://doi.org/10.1109/tpwrs.2007.908471>
- Minsky M (1986) *Society of mind*. Simon and Schuster
- Murray A, Rhymer JEN, Sirmon DG (2021) Humans and technology: forms of conjoined agency in organizations. *Acad Manag Rev* 46(3):552–571
- Park JS, O'Brien J, Cai CJ, Morris MR, Liang P, Bernstein MS (2023) Generative agents: interactive simulacra of human behavior. In: *Proceedings of the 36th annual ACM symposium on user interface software and technology*
- Qin Y, Hu S, Lin Y, Chen W, Ding N, Cui G et al (2025) Tool learning with foundation models. *ACM Comput Surv* 57(4):1–40. <https://doi.org/10.1145/3704435>
- Qu C, Dai S, Wei X, Cai H, Wang S, Yin D et al (2025) Tool learning with large language models: a survey. *Front Comput Sci* 19(8):1–21. <https://doi.org/10.1007/s11704-024-40678-2>
- Queiroz M, Anand A, Baird A (2024) Manager appraisal of artificial intelligence investments. *J Manag Inf Syst* 41(3):682–707
- Reinhard P, Li MM, Peters C, Janson A, Leimeister JM (2026) GenAI-infused service delivery: micro-level augmentation patterns at the service frontline. *J Serv Res*. <https://doi.org/10.1177/10946705251414283>
- Rosemann M, vom Brocke J, Van Looy A, Santoro F (2024) Business process management in the age of AI-three essential drifts. *Inf Syst E-Bus Manag* 22(3):415–429
- Russell S (2019) *Human compatible: AI and the problem of control*. Penguin
- Sapkota R, Roumeliotis KI, Karkee M (2025) AI agents vs. agentic AI: a conceptual taxonomy, applications and challenges. *Inf Fusion*. <https://doi.org/10.1016/j.inffus.2025.103599>
- Saunders C, Benlian A, Henfridsson O, Wiener M (2020) MIS quarterly research curation: IS control & governance. *MIS Q*
- Schneider J, Meske C, Kuss P (2024) Foundation models. *Bus Inf Syst Eng* 66(2):221–231. <https://doi.org/10.1007/s12599-024-00851-0>
- Schneider J (2025) Generative to agentic AI: survey, conceptualization, and challenges. *arXiv*
- Shah MIA, Wahid A, Barrett E, Mason K (2024) Multi-agent systems in peer-to-peer energy trading: a comprehensive survey. *Eng Appl Artif Intell* 132:107847. <https://doi.org/10.1016/j.engappai.2024.107847>
- Shi Y, Xu M, Zhang H, Zi X, Wu Q (2024) A learnable agent collaboration network framework for personalized multimodal AI search engine. In: *Proceedings of the 2nd International Workshop on Deep Multimodal Generation and Retrieval*. ACM, New York, pp 12–20
- Shinn N, Cassano F, Gopinath A, Narasimhan K, Yao S (2023) Reflexion: language agents with verbal reinforcement learning. In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. Curran Associates, Red Hook
- Söllner M, Arnold T, Benlian A, Bretschneider U, Knight C, Ohly S et al (2025) ChatGPT and beyond: exploring the responsible use of generative AI in the workplace: an interdisciplinary perspective. *Bus Inf Syst Eng* 67:289–303
- Steyvers M, Tejeda H, Kumar A, Belem C, Karny S, Hu X et al (2025) What large language models know and what people think they know. *Nat Mach Intell* 7:221–231
- Strong DM, Volkoff O, Johnson SA, Pelletier LR, Tulu B, Bar-On I et al (2014) A theory of organization-EHR affordance actualization. *J Assoc Inf Syst* 15(2):2
- Strunk JA, Banh L, Nissen A, Strobel G, Smolnik S (2024) To delegate or not to delegate? Factors influencing human-agentic IS interaction. In: *ICIS 2024*
- Sujil A, Verma J, Kumar R (2018) Multi agent system: concepts, platforms and applications in power systems. *Artif Intell Rev* 49(2):153–182. <https://doi.org/10.1007/s10462-016-9520-8>
- Tiwana A (2013) *Platform ecosystems: aligning architecture, governance, and strategy*. Morgan Kaufmann, Waltham
- Vu H, Klietsova N, Leopold H, Rinderle-Ma S, Kampik T (2025) Agentic business process management: the past 30 years and practitioners' future perspectives. *arXiv:2504.03693*
- Wang W, Dong L, Cheng H, Liu X, Yan X, Gao J, Wei F (2023) Augmenting language models with long-term memory. *Adv Neural Inf Process Syst* 36:74530–74543
- Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, et al (2022) Chain-of-thought prompting elicits reasoning in large language models. In: *Proceedings of the 36th International Conference on Neural Information Processing Systems*. Curran Associates, Red Hook
- Wessel M, Adam M, Benlian A, Majchrzak A, Thies F (2025) Generative AI and its transformative value for digital platforms. *J Manag Inf Syst* 42(2):346–369
- Wölflein G, Ferber D, Truhn D, Arandjelovic O, Kather JN (2025) Llm agents making agent tools. In: *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*:26092–26130
- Wooldridge M (2009) *An introduction to multiagent systems*. Wiley
- Wooldridge M, Jennings NR (1995) Intelligent agents: theory and practice. *The Knowledge Engineering Review* 10(2):115–152
- Xi Z, Chen W, Guo X, He W, Ding Y, Hong B et al (2025) The rise and potential of large language model based agents: a survey. *Sci China Inf Sci* 68(2):121101
- Yao S, Zhao J, Yu D, Du N, Shafran I, Narasimhan K, Cao Y (eds) (2023a) REACT: synergizing reasoning and acting in language models. In: *11th International Conference on Learning Representations*
- Yu D, Li J, Wang Z, Li X (2024) An overview of swarm coordinated control. *IEEE Trans Artif Intell* 5(5):1918–1938. <https://doi.org/10.1109/TAI.2023.3314581>
- Yue L, Xing S, Chen J, Fu T (2024) ClinicalAgent: clinical trial multi-agent system with large language model-based reasoning. In: *Proceedings of the 15th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, Shenzhen*. ACM, New York
- Shavit Y, Agarwal S, Brundage M, Adler S, O'Keefe C, Campbell R, et al (2023) Practices for governing agentic AI systems. Research Paper, OpenAI. <https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf>. Accessed 26 Apr 2025
- Yang Y, Chai H, Song Y, Qi S, Wen M, Li N, et al (2025) A survey of AI agent protocols. *arXiv:2504.16736*
- Yao S, Yu D, Zhao J, Shafran I, Griffiths TL, Cao Y, Narasimhan K (2023b) Tree of thoughts: deliberate problem solving with large language models. *Curran*. https://proceedings.neurips.cc/paper_files/paper/2023/file/271db9922b8d1f4dd7aaef84ed5ac703-Paper-Conference.pdf
- Zhang J, Lan T, Zhu M, Liu Z, Hoang T, Kokane S, et al (2025) xLAM: a family of large action models to empower AI agent systems. In: *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies*. <https://aclanthology.org/2025.naacl-long.578/>

Zhang J, Hou Y, Xie R, Sun W, McAuley J, Zhao WX, et al (2024) AgentCF: collaborative learning with autonomous language agents for recommender systems. In: Proceedings of the ACM Web Conference 2024. New York. ACM, pp 3679–3689

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.