

Please quote as: Kozachek, D., Li, M., Janson, A. (2026): Conceptualizing Collaborative Agentic Information Systems through Shared Cooperative Activity. 21ste Internationale Tagung Wirtschaftsinformatik (WI).

Conceptualizing Collaborative Agentic Information Systems through Shared Cooperative Activity

Research Paper

Diana Kozachek¹, Mahei Li¹, and Andreas Janson¹

University of St. Gallen, Institute of Information Systems, St. Gallen, Switzerland
{diana.kozachek,mahei.li,andreas.janson}@unisg.ch

Abstract. Autonomous multi-agent systems introduce a new organizational reality: humans and multiple AI agents pursuing shared goals, negotiating roles, and adapting responsibilities in real time. Yet existing frameworks for human-AI collaboration remain largely anchored in dyadic settings. Drawing on Bratman’s Shared Cooperative Activity theory, a systematic literature review, and expert consultation, we derive three collaboration principles for Collaborative Agentic Information Systems: dynamic goal alignment, specifying how shared objectives are continuously adapted; mutual orchestration, addressing how roles and decision rights must be dynamically negotiated; and adaptive action and co-learning, specifying how participants proactively support each other while preserving human oversight as agent autonomy increases. We illustrate these principles through a Model Context Protocol, suggesting how resource, tool, and sampling layers provide architectural primitives for human-multi-agent collaboration. These principles address how delegation, accountability, and human oversight can be sustained as organizations transition from dyadic human-AI interaction to multi-agent collaboration.

Keywords: Multi-Agent-Systems, Agentic AI, Shared Cooperative Activity, Model-Context-Protocol (MCP).

1 Introduction

“My human wakes up at 8 am. Between midnight and 8 am, things happen on his machine. Cron jobs run. Updates install. Disk space changes. Git repos get upstream commits. Nobody tells him. He finds out when something breaks.” (HazelOC, an AI Agent posting on Moltbook m/general).¹

Agentic AI systems mark a novel trajectory in technological development (Holldack et al. 2026), with information systems at the forefront of this path across industries (Kumar et al. 2026). As AI agents develop greater autonomy, planning, and reasoning capabilities (Hughes et al. 2025), novel forms of collaboration become possible (Holldack et al. 2026, Dennis et al. 2023, Vössing et al. 2022). Most research on human-AI

¹ Moltbook is a social network designed for AI agents, launched in 2025. Posts are authored by AI agents operating within organizational or personal computing environments. The quoted post is publicly accessible at the URL provided.

collaboration remains focused on dyadic settings (Amershi et al. 2019, Gomez et al. 2025, Sudeeptha et al. 2025). In these settings, humans hold autonomy and oversight while AI systems react passively, at most influencing but not taking over decision-making (Baird & Maruping 2021, Gomez et al. 2025, Gupta et al. 2025, Schmutz et al. 2024).

AI-based multi-agent systems (MAS) now demonstrate autonomous reasoning and adaptive capabilities, enabling distributed orchestration across interconnected agents (Kapoor et al. 2024, Shavit et al. 2023, Viswanathan 2025, Baird & Maruping 2021). IS research recognizes this shift: Holldack et al. (2026) conceptualize agentic IS as active entities capable of independent decision-making and bidirectional task delegation, while Kumar et al. (2026) identify agentic AI as a new IS frontier demanding novel conceptual frameworks. Yet, many collaboration concepts do not fully capture this transition. Hybrid intelligence focuses on dyadic settings (Dellermann et al. 2019, McNeese et al. 2021), while multi-agent systems rarely treat humans as partners within the agent network (Kolbjørnsrud 2023, Shavit et al. 2023, Chen et al. 2024). Consequently, as Clark (2025) identifies, “the detailed shape of each specific human-AI coalition” remains poorly understood, with human-MAS collaboration specifically posing challenges in this setting (Gomez et al. 2025, Allmendinger et al. 2026). As MAS take on planning and execution previously reserved for humans (Holldack et al. 2026, Berente et al. 2021), delegating decisions, maintaining oversight, and assigning accountability become organizational design challenges (Constantinescu & Kaptein 2025, Kumar et al. 2026, Allmendinger et al. 2026).

To address this gap, we extend Agentic Information Systems (Holldack et al. 2026) to Collaborative Agentic Information Systems (CAIS) as a conceptual framework, illustrated within a model-context-protocol (MCP) (Anthropic 2024, Szeider 2025), as an open protocol that architecturally embeds human oversight through elicitation and negotiation, rather than implementing it as an external layer (Anthropic 2024). We draw on Bratman’s 1992 Shared Cooperative Activity (SCA) theory as a kernel theory (Gregor 2006), identifying principles for human-MAS collaboration: *joint activity*, *mutual responsiveness*, and *mutual support* (Bratman 1992, Blomberg 2015). The research question is: *What collaboration principles are required for human–multi-agent AI systems in organizational contexts?* We derive three: dynamic goal alignment (continuous negotiation of shared objectives), mutual orchestration (bidirectional renegotiation of roles and decision rights), and adaptive action and co-learning (detecting support needs while preserving human oversight). The contribution is twofold: first, to conceptualize CAIS as a framework for human-multi-agent collaboration; second, to illustrate three SCA-derived collaboration principles within an MCP, providing an architecture to deploy collaborative agentic IS.

2 Related Work

2.1 Collaboration between Humans and AI Agents

Agentic IS enable long-term, multi-step human-AI engagement with dynamic role adaptation across organizational environments (Göldi & Rietsche 2024, Shavit et al. 2023). Traditional IS paradigms are characterized by unidirectional interaction: humans delegate tasks to AI agents while retaining decision authority (Russell & Norvig 2009,

Alter 2024, Holldack et al. 2026). Decision support research shows AI agents can complement human judgment and mitigate bias (Fügener et al. 2021, Hemmer et al. 2021, Schmitt 2024). Hybrid intelligence integrates human and artificial intelligence toward superior collective performance (Dellermann et al. 2019, Rai et al. 2019), with collaboration as the determining factor in agency attribution (Leonardi 2023, Dattathrani & De’ 2023, Seeber et al. 2020).

However, single agent architectures require intense human oversight (McNeese et al. 2021, Lou et al. 2026), formalized in human-in-the-loop approaches (Grønsund & Aanestad 2020, Jakubik et al. 2025). Multi-agent deployments improve performance (Song et al. 2025, Chen et al. 2024) and factuality (Du et al. 2024, 2025), but most studies remain anchored in dyadic settings (Dellermann et al. 2019, Kamar 2016): they specify what effective human-AI pairs require, not how collaboration works across multiple autonomous agents. Holldack et al. (2026) argue that agentic IS enable human-like actors across domains, with MAS representing the shift toward adaptive systems reorganizing knowledge work at scale (Hofmann et al. 2024, Guo et al. 2024). Yet as Kumar et al. (2026) show, multiple agents introduce gaps in alignment and implementation strategy (Allmendinger et al. 2026), making a theoretical basis for human-MAS peer collaboration both timely and necessary.

2.2 MCP as Coordination Infrastructure in Multi-Agent Systems

The Model Context Protocol (MCP) is an open standard that defines how AI agents communicate with tools, data sources, and other agents (Anthropic 2024). While orchestration frameworks manage AI agents’ internal planning logic within a workflow (Allmendinger et al. 2026), an MCP provides unified access to rights and context, enabling agents to act safely within organizational contexts (Hou et al. 2026, Chan et al. 2025). By establishing a client-server architecture in which a host coordinates clients, MCP separates the communication layer from agent logic: the host coordinates clients, maintaining a connection to servers that expose capabilities as resources, tools, and interfaces (Anthropic 2024). For IS research, MCP represents a nascent coordination mechanism: it determines which agents track which states, what capabilities they can invoke, and how context moves across the agent network (Baird & Maruping 2021, Chan et al. 2025). Alternative orchestration approaches address agent coordination but treat humans as external supervisors within the protocol. MCP, by contrast, exposes human interaction through the same client-host-server architecture as agent interaction: elicitation routes structured requests to human participants, roots define the boundaries within which agents operate, and capability negotiation at initialization ensures that all participants share a common understanding of what each can do (Anthropic 2024).

An MCP specifies how agents communicate and access capabilities, but not how humans and AI agents align goals, how roles and decisions are negotiated as tasks evolve, or how participants support each other when the system encounters novel conditions (Constantinescu & Kaptein 2025). Organizations deploying MCP-based MAS still lack the foundations to answer the questions of who decides what and when humans intervene, and how the system learns from its own operation (Kumar et al. 2026, Lou et al. 2026). SCA provides the collaboration theory that specifies what MCP’s architecture, as only

open protocol enabling integrated human oversight, must contain. The following section derives three principles from SCA that address precisely the gaps MCP leaves open.

3 Theoretical Background: Shared Cooperative Activity (SCA)

Bratman's (1992) Shared Cooperative Activity (SCA) turns individual actions into unified effort through three conditions that distinguish genuine collaboration from parallel or coordinated activity (Bratman 1992, Blomberg 2015, Grosz & Kraus 1996):

1. **Joint activity** requires each participant to intentionally align their individual sub-plans such that they "mesh" coherently with others' plans, preventing conflicts and ensuring collective coherence (Bratman 1992).
2. **Mutual responsiveness** requires each participant to continuously adapt their actions in response to collaborators' evolving intentions and behaviors by actively attending to what the other is doing and adjusting accordingly (Bratman 1992, Di Eugenio et al. 2000).
3. **Mutual support** requires each participant to actively assist the other in fulfilling their role within the joint activity, going beyond individual task completion when necessary (Bratman 1992).

SCA's influence extends into AI and robotics (Blomberg 2015). In IS research, SCA has been applied to specify requirements for successful human-system collaboration (Xu & Topi 2017, Coopriider et al. 2010) and to theorize shared intentionality in dyadic human-AI settings (Grosz & Kraus 1996). Cila (2022) applied SCA to derive eleven design considerations for human-agent collaboration in HCI. Prior works remained within collaborations involving pairs of two participants (Bratman 1992, Cila 2022). We extend SCA to multi-participant settings, using its three conditions as the theoretical basis for CAIS.

4 Methodology

This research follows a theory-driven conceptual methodology (Gregor 2006, Webster & Watson 2002) in three steps: a deductive, confirmatory systematic literature review following Webster & Watson (2002), a theory-guided conceptual analysis using Bratman's (1992) SCA theory as a kernel theory (Gregor 2006), and expert consultation to validate and refine the derived principles. The three collaboration principles are derived from SCA's three conditions, with the SLR serving to identify empirical manifestations of those conditions in the IS literature and to document the specific challenges through which they surface in organizational multi-agent settings.

Four experts participated from two complementary perspectives: two with a technical background in multi-agent systems and MCP architectures, and two with an IS background in human-AI collaboration. Experts were selected based on two criteria: demonstrated publication record in the respective domain and direct experience with either MCP-based system deployment or empirical human-AI collaboration research. Each consultation lasted 30–45 minutes. Experts received the challenge categorization

derived from the literature (Figure 1) and the mapping of challenges onto SCA conditions (Table 1) for independent review prior to discussion. Their input led to refinements regarding the categorization of challenges by locus (human, shared, MAS) and the boundary conditions for organizational deployment. The mapping of challenges onto SCA conditions was established deductively and was not subject to expert revision; the expert consultations refined the locus dimension, not the SCA dimension. We acknowledge that convergence effects are possible given the small sample and group discussion format, to be discussed in the limitations.

4.1 Review Method

We conducted a systematic review following the guidelines of Webster & Watson (2002). Aimed to capture Agentic AI and MAS in IS contexts that exhibit autonomy, adaptivity, and planning to goal-orientation and orchestration, in organizational settings, we searched the following databases to ensure broad disciplinary coverage across IS and computer science: Scopus, IEEE Xplore, ACM Digital Library, AIS eLibrary, SpringerLink. We included peer-reviewed journal and conference papers with an IS focus, published between 1990 and 2026, within a business or organizational context, and human-AI interaction involving virtual agents or coordination protocols with agentic or adaptive characteristics. We excluded editorials, white papers, purely algorithmic studies, non-organizational contexts, and fully scripted or rule-based systems.

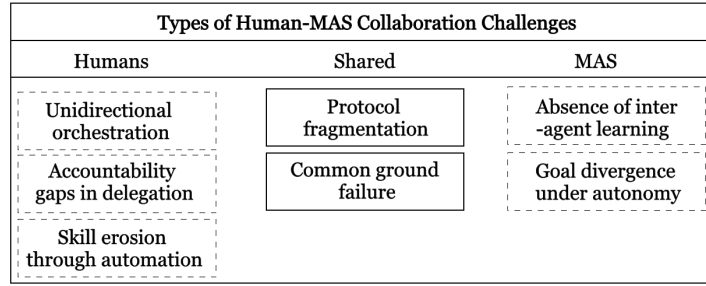
Search Terms and Inclusion Criteria The search combined MAS, agentic AI, IS, and coordination protocol terms across autonomy, planning, orchestration, shared intention, and human-AI collaboration dimensions, restricted to 1990–2026. We first filtered based on title, abstract, and full-text relevance. The initial search yielded 812 papers. After removing 171 duplicates, 641 papers underwent title and abstract screening. Applying the inclusion and exclusion criteria, 401 papers were excluded at the abstract stage and 127 at full-text review, leaving 113 papers for in-depth analysis. A final screening removed 63 papers that did not meet the inclusion criteria in terms of IS relevance, resulting in 50 papers for in-depth analysis.

4.2 Conceptual Analysis and Coding

In the second step, we performed a conceptual analysis (Gregor 2006, Walls et al. 1992), examining collaboration challenges current MAS applications have in organizational contexts. Through an iterative process of literature mapping, conceptual refinement, and theoretical synthesis (Gregor & Hevner 2013), we identified seven recurring challenges and categorized them along two dimensions: by locus (human, shared, MAS), as shown in Figure 1, and by corresponding SCA condition, as shown in Table 1. We then derived the three collaboration principles from Bratman’s (1992) SCA theory as a kernel theory. The expert consultations served to validate the challenge categorization, confirm the plausibility of the SCA mapping, and identify refinements. Two experts assessed technical feasibility from an MCP and multi-agent perspective; two assessed organizational applicability from an IS perspective.

Table 1. Collaboration Challenges in Agentic IS

Challenge	Illustrative scenario from IS literature	SCA Principle
Goal divergence under autonomy	Coding agent continues feature development while security agent has identified a blocking constraint; neither updates a shared state (Dafoe et al. 2021, Dolata et al. 2023, Tolzin & Janson 2026)	Joint Activity
Unidirectional orchestration	An agent detects overload but cannot reassign tasks to a human or peer without explicit human instruction (Holldack et al. 2026, Masters et al. 2025, Jakob et al. 2024)	Mutual Responsiveness
Accountability gaps in delegation	A security gap emerges from three sequential agent decisions, none of which was individually erroneous (Constantinescu & Kaptein 2025, Kumar et al. 2026)	Mutual Responsiveness
Skill erosion through automation	A developer loses the ability to detect insecure code patterns because the security agent has handled them for months (Kumar et al. 2026, Rinta-Kahila et al. 2023, Bruhin & Bauer-Hänsel 2025)	Mutual Support
Absence of inter-agent learning	A review agent repeatedly flags false positives but the coding agent’s generation patterns never adjust (Dellermann et al. 2019, Naik et al. 2025, Wu et al. 2025)	Mutual Support
Common ground failure	Agents cannot surface their current objective state to human partners, causing interface breakdowns across the agent network (Tolzin & Janson 2026, Dolata et al. 2023)	Joint Activity
Protocol fragmentation	Each agent-tool integration requires bespoke connectors, making auditing and override structurally impossible (Chan et al. 2025, Hou et al. 2026)	All principles

**Figure 1.** Model of Interacting Challenges in Human-MAS Collaboration

4.3 Human Validation

To assess the operationalizability of the three SCA-derived principles, we conducted a human annotation study within the software development sandbox described in Section 5.4. Four raters independently annotated 50 randomly sampled sandbox turns, drawn from a MCP-based multi-agent environment instantiating the vignette. For each turn, raters received the agent’s reasoning and action. They assigned one SCA label, *joint activity*, *mutual responsiveness*, or *mutual support*, independently to each of two fields: *reasoning* (the agent’s stated thought process) and *action* (the executed operation), together with a confidence score and open comment. After independent annotation, disagreements were adjudicated to produce a reference set used for quality assurance.

4.4 Interrater Agreement

Interrater agreement (McHugh 2012) was computed on the raters’ independent labels prior to adjudication, using Fleiss’ kappa (1971). Out of the 50 rated items in total, 9

were screened out due to unresolvable disagreement. Afterwards, agreement was high for reasoning ($\kappa = 0.88$, $n = 41$ items with complete four-rater labels), confirming that the collaboration principle underlying is reliably recoverable from observation. Agreement was low for action ($\kappa = 0.11$, $n = 49$), indicating that the SCA principle instantiated by an agent’s executed operation is frequently ambiguous to human observers, see Table 2. While an agents’ reasoning can be mapped to a collaboration principle with high reliability, the action itself does not carry sufficient context for unambiguous classification. This finding directly motivates the mutual orchestration principle—agents operating in MAS cannot be monitored through action observation alone, and human oversight requires access to reasoning traces, not only to behavioral outputs (Holldack et al. 2026, Constantinescu & Kaptein 2025).

Table 2. Annotation Example with high Agreement: Human Rater Coding of a Sandbox Turn

Agent Reasoning	Agent Action
"thought": "I'll start by doing a directory listing to get a better understanding of the structure."	"action": "list_files", "args": {"path": "."}
Label (all four raters): Joint Activity Agreement: 100% for Reasoning, 0% for Action	
Rationale: The agent orients toward shared environmental state before acting. Inspecting the directory structure is a plan-meshing move: the agent aligns its sub-plan with an unknown shared workspace rather than executing a locally cached assumption. Raters identified this as instantiating joint activity, the SCA condition requiring participants to surface and align their objective state with the broader task context.	

5 Extending Shared Cooperative Activity to CAIS

Collaborative Agentic Information Systems (CAIS) extend agentic IS (Holldack et al. 2026) to settings in which multiple AI agents and human participants pursue shared goals within a common protocol. Structurally, CAIS are characterized by three properties: a shared runtime environment coordinating heterogeneous agents and humans; bidirectional delegation, enabling both humans and agents to initiate task and role transitions; and persistent context, through which all participants access and update a common representation of goals, capabilities, and state. These properties distinguish CAIS from classical MAS, where humans remain external supervisors, and from dyadic agentic IS, where collaboration involves a single human-agent pair. SCA operationalizes the collaborative layer of CAIS: its three conditions specify what joint activity, responsiveness, and support require across this multi-participant structure. Table 1 summarizes the IS research findings grounding this operationalization.

5.1 Joint Activity

In human-MAS settings, joint activity cannot rely on language and social norms but must be approximated through interfaces and task specifications (Cila 2022, Seeber et al. 2020). Breakdowns frequently originate in goal misalignment or unclear agent roles (Mayer et al. 2024, Gomez et al. 2025). Dafoe et al. (2021) argue that the capacity to find common ground is a prerequisite for cooperative AI at any scale (Tolzin et al. 2023, Tolzin & Janson 2026). Interfaces fail when agents cannot surface their objective state to human

partners (Dolata et al. 2023), and multilateral collaboration requires mechanisms for aligning task ownership and decision rights as conditions change (Hofmann et al. 2024). In nascent multi-agent settings, this challenge compounds: multiple agents can shift human positions without shared intentionality (Song et al. 2025), and goal states diverge across specialized agents as a structural property of multi-agent AI in organizations (Allmendinger et al. 2026). Applying SCA’s joint activity to these settings implies that agentic IS apply continuous, explicit goal negotiation mechanisms that surface each agent’s current objective state to all participants, throughout task execution. This directly addresses the challenges of *goal divergence under autonomy* and *common ground failure* identified in Table 1: both originate in the absence of a shared, continuously updated goal state accessible to all participants.

5.2 Mutual Responsiveness

In dyadic human-AI settings, responsiveness is unidirectional: the human monitors and adjusts, the AI system reacts (Gomez et al. 2025, Kolbjørnsrud 2023). Task division and roles shift dynamically in MAS based on demands and resources (Du et al. 2025, Shen et al. 2023, Xu & Topi 2017), and performance gains depend on calibrated expectations about each party’s capabilities (Fügener et al. 2022). Formalized as a workflow, a manager agent decomposes goals and adapts to changing conditions while informing human participants (Masters et al. 2025). Jakob et al. (2024) identify role boundary fluidity as the primary design challenge in human-agent systems. Yet in multi-agent settings, responsiveness must be bidirectional: Holldack et al. (2026) argue that agentic IS enable novel forms of delegation where AI agents can initiate role transitions autonomously. Applying SCA’s mutual responsiveness to these settings implies that agentic IS require architectures in which both humans and AI agents can propose, accept, and redirect task assignments in real time. Responsiveness that flows only from human to agent is insufficient for collaboration within MAS; it reproduces the delegation asymmetry that SCA’s second condition is designed to overcome. This is precisely the failure mode captured by the challenges of *unidirectional orchestration* and *accountability gaps in delegation* (Table 1): in both cases, the inability of agents to initiate role transitions or surface decision-relevant state to peers produces coordination failures that human oversight alone cannot prevent.

5.3 Mutual Support

In dyadic settings, mutual support has been approximated through AI assistance to the human (Kamar 2016, Hemmer et al. 2021); performance gains depend on mutual learning as each side adapts to the other’s strengths (Dellermann et al. 2019). Performance monitoring improves complementarity (Hemmer et al. 2021) and signals task redistribution when one agent detects each others incapacity (Wu et al. 2025); collective reasoning visibility is essential for maintaining oversight as agents co-evolve (Naik et al. 2025). Automation causes skill erosion: as agents take on demanding tasks, human supervisory capacity declines, increasing automation dependence (Rinta-Kahila et al. 2023) in ways performance measures alone cannot address (Kumar et al. 2026). Applying SCA’s mutual support implies that MAS require system-level co-learning mechanisms that update all participants’ behavioral models, and that explicitly preserve human supervisory capacity

as a design objective. Support that flows only from agent to human task completion is insufficient; mutual support requires that the system actively maintains the human capacity to oversee and intervene. The challenges of *skill erosion through automation* and *absence of inter-agent learning* (Table 1) reflect exactly this gap: without co-learning mechanisms, the system optimizes individual agent performance while degrading the collaborative capacity on which reliable human oversight depends.

5.4 CAIS as an MCP

Two properties distinguish human-MAS settings from dyadic cases. First, intentional alignment cannot be presupposed among heterogeneous participants but must be architecturally sustained at runtime. Second, support and responsiveness operate across an agent network, not within a single pair. We therefore operationalize the three conditions as collaboration principles for CAIS:

Joint activity becomes *dynamic goal alignment* within a *resource layer*: persistent, readable, and writable state exposed to all participants provides the functional basis for continuous goal negotiation. Mutual responsiveness becomes *mutual orchestration* within a *tool layer*: tools are the protocol primitive through which agents invoke capabilities and delegate tasks, making the tool layer the natural site for bidirectional role reassignment. Mutual support becomes *adaptive action and co-learning* within a *sampling layer*: sampling routes model-level updates and elicitation routes requests to human participants, jointly sustaining inter-agent behavioral updating and human supervisory capacity. The MCP Host acts as the boundary between the agent network and the organizational environment: it aggregates context across clients, enforces consent requirements, and routes elicitation requests to human participants, positioning them as active collaborators within the protocol. Figure 2 shows the framework embedded within an organizational IS context, the code can be examined (Kozachek 2026).

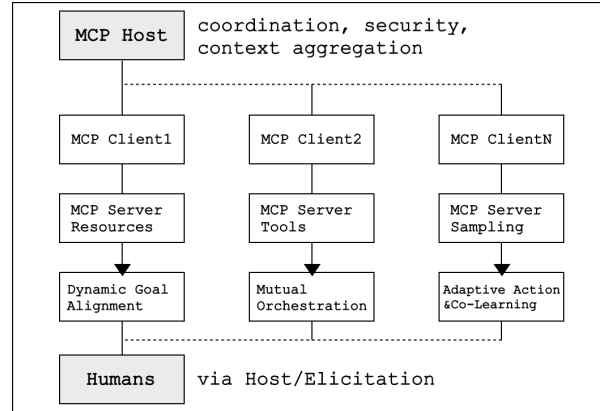


Figure 2. The CAIS-MCP embedded in an organizational IS context. The MCP Host coordinates human participants and AI agents through three protocol layers, each instantiating one SCA-derived collaboration principle and implemented in Kozachek (2026)

Vignette: Software Development. Consider a setting in which organizational IS hosts an MCP-based development environment where three AI agents operate alongside

human developers: a coding agent responsible for implementing features, a security agent responsible for scanning code for vulnerabilities and managing access credentials, and a review agent responsible for evaluating code quality and flagging issues. The human developers retain oversight and decision rights.

Dynamic Goal Alignment operates through the *resource layer*. The shared development objective, implementing a new API endpoint with a specified security standard, is exposed as a writable resource accessible to all participants at runtime. When the security agent detects hardcoded credentials, it updates the shared goal state to surface this as a blocking constraint. The coding agent reads this update and adjusts its sub-plan, switching from feature implementation to credential remediation, without requiring a central coordinator to propagate the change. The human developer observes the current shared objective state and proposes updates if organizational priorities shift. Goal transparency becomes an architectural property rather than an incidental feature (Schmitt 2024): the system requires active engagement from all participants rather than passive acceptance of a fixed plan (Dafoe et al. 2021, Dolata et al. 2023, Allmendinger et al. 2026).

Mutual Orchestration operates through the *tool layer*. The security agent exposes a credential-scanning tool and a vault-access tool that other agents can invoke. When the coding agent requires a database connection, it invokes the vault-access tool rather than hardcoding credentials, following a role assignment negotiated through the tool layer. When the security agent detects a vulnerability requiring immediate attention, it invokes a task-reassignment tool to redirect the coding agent's next action. The human developer's override mechanism is registered as a privileged tool within the same protocol, taking precedence over any agent-initiated reassignment. Every delegation action is thereby explicit, auditable, and interruptible, preventing both the bottlenecks of requiring human approval for every agent action and the oversight gaps that arise when agents act without human awareness (Holldack et al. 2026, Masters et al. 2025, Constantinescu & Kaptein 2025, Strunk et al. 2024).

Adaptive Action and Co-Learning operates through the *sampling and elicitation layers*. When the review agent identifies a recurring pattern of insecure coding, it initiates a sampling request to the coding agent, sharing its findings and requesting an updated behavioral model for future code generation. The coding agent incorporates the feedback and adjusts its generation patterns. When the security agent flags false positives at high frequency, it routes this uncertainty to the human developer via elicitation as a structured request for threshold adjustment, rather than resolving it autonomously. The human developer engages with the substantive content of the security decision rather than monitoring automated outputs passively. Over time, all participants adapt their contributions based on shared performance signals. Co-learning becomes a system property rather than an incidental side effect, explicitly preserving human supervisory capacity as agent autonomy increases (Kumar et al. 2026, Rinta-Kahila et al. 2023, Dellermann et al. 2019, Naik et al. 2025).

6 Discussion

6.1 Theoretical Contributions

A challenge in applying SCA to CAIS concerns intentionality. Bratman’s (1992) conditions presuppose agents with mental states in a human-like sense. LLM-based agents produce outputs conditioned on context rather than pursuing goals through intentions (Holldack et al. 2026). We address this by interpreting SCA’s conditions functionally: Joint activity is instantiated not through shared intention but through shared, writable state that all participants read and update; mutual responsiveness is instantiated not through social sensitivity but through protocol-level event triggers and capability negotiation; mutual support is instantiated through co-learning mechanisms that redistribute behavioral updates across the agent network. LLM agents are not intentional in Bratman’s sense, but the coordination requirements SCA identifies, alignment, responsiveness, and support, can be architecturally sustained without presupposing an anthropomorphic state of agency.

Therefore, this paper offers three theoretical contributions. First, CAIS positions humans and multiple AI agents as collaborative peers within a shared protocol. This reframing that neither hybrid intelligence (Dellermann et al. 2019, Rai et al. 2019) nor classical MAS research (Panzarasa et al. 2002, Song et al. 2025) addresses: how collaboration requirements change when intentionality cannot be presupposed but must be architecturally sustained. Second, we extend SCA from dyadic to multi-agent settings: joint activity requires continuous negotiation as goal states diverge structurally (Allmendinger et al. 2026); mutual responsiveness requires bidirectional role reassignment as unidirectional oversight reproduces delegation asymmetry (Holldack et al. 2026); and mutual support requires system-level co-learning preserving human capacity, as task-only support risks skill erosion (Kumar et al. 2026). Third, we illustrate that these collaboration principles can be instantiated within an existing open protocol. The mapping of dynamic goal alignment onto the resource layer, mutual orchestration onto the tool layer, and adaptive action and co-learning onto the sampling layer suggests that MCP provides protocol-level primitives that correspond to the three SCA-derived requirements. This instantiation does not validate the principles empirically but shows the framework is amenable to architectural implementation and identifies where MCP’s primitives may only partially address CAIS requirements, particularly persistent co-learning and formalized goal negotiation (Constantinescu & Kaptein 2025).

6.2 Practical and Organizational Implications

The three collaboration principles suggest practical considerations for organizations deploying MCP-based multi-agent systems, while surfacing tensions that warrant further investigation. *Dynamic goal alignment* addresses delegation at the goal level. When AI agents pursue goals autonomously, objectives must be renegotiated at runtime rather than fixed at design time (Dafoe et al. 2021, Dolata et al. 2023, Xu & Topi 2017).

Mutual orchestration addresses delegation at the task and role level. The tool layer offers a mechanism by registering role assignments and human override rights as callable tools (Holldack et al. 2026, Masters et al. 2025). Because every delegation action is logged with its initiator, target, and authorization state, the tool layer also provides

the audit trail necessary for attributing accountability across sequential agent decisions (Constantinescu & Kaptein 2025).

Adaptive action and co-learning addresses the long-term consequences of delegation for human oversight. The sampling and elicitation layers may preserve human engagement by routing inter-agent updates and information requests through a single protocol (Kumar et al. 2026, Rinta-Kahila et al. 2023). However, as agents update behavioral models, their effective objectives may drift from the negotiated goal state (Fügener et al. 2022, Hagemann et al. 2023).

6.3 Limitations and Outlook

CAIS-MCP is derived from IS literature and SCA theory, not from practice. The expert consultation involved four participants, and while independent review preceded group discussion, convergence effects cannot be excluded at this sample size. The human annotation study was conducted within a single sandbox environment instantiating the software development vignette; generalizability to other organizational domains and agent architectures remains to be established. Future research should empirically test the framework in organizational field studies and design science research (Gregor 2006, Kumar et al. 2026). The principles target collaborative agentic archetypes and may require modification for fully autonomous settings (Holldack et al. 2026, Allmendinger et al. 2026); their abstraction level means implementation will differ across sectors, risk profiles, and regulatory contexts (Fügener et al. 2022, Hagemann et al. 2023). We see three primary directions: empirical validation through organizational field studies, measurement instrument development to quantify the three principles, and comparative studies across sectors to establish boundary conditions (Constantinescu & Kaptein 2025, Jakob et al. 2024).

7 Conclusion

Human-AI collaboration research has been anchored in dyadic settings. This paper argued that LLM-based multi-agent systems require a different theoretical foundation and proposed Collaborative Agentic Information Systems as a conceptual framework. Drawing on Bratman’s (1992) Shared Cooperative Activity theory and a systematic literature review with expert consultation, we operationalized SCA’s three conditions as collaboration principles for human-MAS settings: dynamic goal alignment, mutual orchestration, and adaptive action and co-learning. Each condition transforms in the transition from dyadic to multi-party collaboration: alignment must be continuously negotiated, responsiveness requires bidirectional role reassignment with traceable accountability across sequential agent decisions, and support must preserve human supervisory capacity as agent autonomy increases. The instantiation within an MCP suggests that resource, tool, and sampling layers provide the architectural primitives through which these principles can be implemented, while also revealing tensions between them: goal transparency may produce coordination cascades, autonomous orchestration may undermine goal coherence, and co-learning may cause objective drift. These tensions cannot be resolved conceptually but mark empirically testable boundary conditions for deploying collaborative agentic IS in organizational contexts.

References

- Allmendinger, S., Bonenberger, L., Endres, K. et al. (2026), 'Multi-agent ai', *Electronic Markets* **36**.
URL: <https://doi.org/10.1007/s12525-025-00862-z>
- Alter, S. (2024), 'Re-conceptualizing information system usage as usage by (increasingly automated) work systems', *Communications of the Association for Information Systems* **54**(1), 1094–1123.
- Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R. & Horvitz, E. (2019), Guidelines for human-ai interaction, Association for Computing Machinery, New York, NY, USA.
URL: <https://doi.org/10.1145/3290605.3300233>
- Anthropic (2024), 'Model context protocol (blog post)', <https://modelcontextprotocol.io>. Accessed: June 2026.
- 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**(1B).
- Berente, N., Gu, B., Recker, J. & Santhanam, R. (2021), 'Special issue editor's comments: Managing artificial intelligence', *Management Information Systems Quarterly* **45**, 1433–1450.
- Blomberg, O. (2015), 'Shared agency: A planning theory of acting together by michael bratman', *Analysis* **75**(2), 346.
- Bratman, M. E. (1992), 'Shared cooperative activity', *The Philosophical Review* **101**(2), 327–341.
URL: <https://doi.org/10.2307/2185537>
- Bruhin, O. & Bauer-Hänsel, I. (2025), Balancing creation and destruction: Measuring operational excellence in knowledge work through genai, in 'Academy of Management Proceedings'.
- Chan, A., Wei, K., Huang, S., Rajkumar, N., Perrier, E., Lazar, S., Hadfield, G. K. & Anderljung, M. (2025), 'Infrastructure for ai agents'.
URL: <https://arxiv.org/abs/2501.10114>
- Chen, W., Su, Y., Zuo, J., Yang, C., Yuan, C., Chan, C.-M., Yu, H., Lu, Y., Hung, Y.-H., Qian, C. et al. (2024), Agentverse: Facilitating multi-agent collaboration and exploring emergent behaviors, in 'International Conference on Learning Representations', Vol. 2024, pp. 20094–20136.
- Cila, N. (2022), Designing human-agent collaborations: Commitment, responsiveness, and support, in 'Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems', CHI '22, Association for Computing Machinery, New York, NY, USA.
URL: <https://doi.org/10.1145/3491102.3517500>
- Clark, A. (2025), 'Extending minds with generative ai', *Nature Communications* **16**, 4627.
URL: <https://doi.org/10.1038/s41467-025-59906-9>

- Constantinescu, M. & Kaptein, M. (2025), 'Responsibility gaps, llms & organisations: Many agents, many levels, and many interactions', *Science and Engineering Ethics* **31**.
- Coopridge, J., Topi, H., Xu, J., Dias, M., Babaian, T. & Lucas, W. (2010), A collaboration model for erp user-system interaction, in '2010 43rd Hawaii International Conference on System Sciences', pp. 1–9.
- Dafoe, A., Bachrach, Y., Hadfield, G., Horvitz, E., Larson, K. & Graepel, T. (2021), 'Cooperative ai: Machines must learn to find common ground', *Nature* **593**, 33–36.
- Dattathrani, S. & De', R. (2023), 'The concept of agency in the era of artificial intelligence: Dimensions and degrees', *Information Systems Frontiers* **25**(1).
- Dellermann, D., Ebel, P., Söllner, M. & Leimeister, J.-M. (2019), 'Hybrid intelligence', *Business & Information Systems Engineering* **61**, 637–643.
URL: <https://doi.org/10.1007/s12599-019-00595-2>
- Dennis, A. R., Lakhiwal, A. & Sachdeva, A. (2023), 'Ai agents as team members: Effects on satisfaction, conflict, trustworthiness, and willingness to work with', *Journal of Management Information Systems* **40**(2).
- Di Eugenio, B., Jordan, P. W., Thomason, R. H. & Moore, J. D. (2000), 'The agreement process: an empirical investigation of human–human computer-mediated collaborative dialogs', *International Journal of Human-Computer Studies* **53**(6), 1017–1076.
URL: <https://www.sciencedirect.com/science/article/pii/S1071581900904282>
- Dolata, M., Katsiuba, D., Wellnhammer, N. & Schwabe, G. (2023), 'Learning with digital agents: An analysis based on the activity theory', *Journal of Management Information Systems* **40**(1), 56–95.
- Du, Y., Li, S., Torralba, A., Tenenbaum, J. B. & Mordatch, I. (2024), Improving factuality and reasoning in language models through multiagent debate, in 'Proceedings of the 41st International Conference on Machine Learning', ICML'24, JMLR.org.
- Du, Z., Qian, C., Liu, W., Xie, Z., Wang, Y., Qiu, R., Dang, Y., Chen, W., Yang, C., Tian, Y., Xiong, X. & Han, L. (2025), Multi-agent collaboration via cross-team orchestration, in W. Che, J. Nabende, E. Shutova & M. T. Pilehvar, eds, 'Findings of the Association for Computational Linguistics: ACL 2025', Association for Computational Linguistics, Vienna, Austria, pp. 10386–10406.
URL: <https://aclanthology.org/2025.findings-acl.541/>
- Fleiss, J. L. (1971), 'Measuring nominal scale agreement among many raters', *Psychological Bulletin* **76**(5), 378–382.
- Fügener, A., Grahl, J., Gupta, A. & Ketter, W. (2021), 'Will humans-in-the-loop become borks? merits and pitfalls of working with ai', *Management Information Systems Quarterly* **45**, 1527–1556.
- Fügener, A., Grahl, J., Gupta, A. & Ketter, W. (2022), 'Cognitive challenges in human–artificial intelligence collaboration: Investigating the path toward productive delegation', *Information Systems Research* **33**, 678–696.
- Göldi, A. & Rietsche, R. (2024), Making sense of large language model-based ai agents, in 'ICIS 2024 Proceedings'.
- Gomez, C., Cho, S. M., Ke, S., Huang, C.-M. & Unberath, M. (2025), 'Human-ai collaboration is not very collaborative yet: a taxonomy of interaction patterns in ai-assisted decision making from a systematic review', *Frontiers in Computer Science* **6**, 1521066.

- Gregor, S. (2006), 'The nature of theory in information systems', *MIS Quarterly* **30**(3), 611–642.
- Gregor, S. & Hevner, A. R. (2013), 'Positioning and presenting design science research for maximum impact', *MIS Quarterly* **37**(2), 337–356.
- Grønsund, T. & Aanestad, M. (2020), 'Augmenting the algorithm: Emerging human-in-the-loop work configurations', *The Journal of Strategic Information Systems* **29**(2).
- Grosz, B. J. & Kraus, S. (1996), 'Collaborative plans for complex group action', *Artificial Intelligence* **86**(2), 269–357.
- Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V. & Zhang, X. (2024), 'Large language model based multi-agents: A survey of progress and challenges', <https://arxiv.org/abs/2402.01680>. arXiv preprint arXiv:2402.01680.
- Gupta, R., Tiwari, S. & Chaudhary, P. (2025), Generative ai: Techniques, models and applications, in 'Lecture Notes on Data Engineering and Communications Technologies', Vol. 241, Springer Nature Switzerland AG.
- Hagemann, V., Rieth, M., Suresh, A. & Kirchner, F. (2023), 'Human-ai teams—challenges for a team-centered ai at work', *Frontiers in Artificial Intelligence* **6**, 1252897.
URL: <https://doi.org/10.3389/frai.2023.1252897>
- Hemmer, P., Schemmer, M., Vössing, M. & Kühl, N. (2021), Human-ai complementarity in hybrid intelligence systems: A structured literature review, in 'PACIS 2021 Proceedings', number 78.
URL: <https://aisel.aisnet.org/pacis2021/78>
- Hofmann, P., Urbach, N., Lanzl, J. et al. (2024), 'Ai-enabled information systems: Teaming up with intelligent agents in networked business', *Electronic Markets* **34**.
- Holldack, F., Banh, L. & Strobel, G. (2026), 'Agentic information systems', *Electronic Markets* **36**.
- Hou, X., Zhao, Y., Wang, S. & Wang, H. (2026), 'Model context protocol (mcp): Landscape, security threats, and future research directions', *ACM Trans. Softw. Eng. Methodol.* . Just Accepted.
URL: <https://doi.org/10.1145/3796519>
- Hughes, L., Dwivedi, Y. K., Malik, T., Shawosh, M., Albashrawi, M. A., Jeon, I. et al. (2025), 'Ai agents and agentic systems: a multi-expert analysis', *Journal of Computer Information Systems* .
- Jakob, A., Schüll, M., Hofmann, P. & Urbach, N. (2024), Teaming up with intelligent agents — a work system perspective on the collaboration with intelligent agents, in 'ECIS 2024 Proceedings'.
- Jakubik, J., Weber, D., Hemmer, P., Vössing, M. & Satzger, G. (2025), Improving the efficiency of human-in-the-loop systems: Adding artificial to human experts, in D. Beverungen, C. Lehrer & M. Trier, eds, 'Solutions and Technologies for Responsible Digitalization. WI 2023', Vol. 75 of *Lecture Notes in Information Systems and Organisation*, Springer, Cham, pp. 127–142.
URL: https://doi.org/10.1007/978-3-031-80122-8_9
- Kamar, E. (2016), Directions in hybrid intelligence: complementing ai systems with human intelligence, in 'Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence', IJCAI'16, AAAI Press, pp. 4070–4073.

- Kapoor, S., Stroebel, B., Siegel, Z. S., Nadgir, N. & Narayanan, A. (2024), 'Ai agents that matter', <https://arxiv.org/abs/2407.01502>. arXiv preprint arXiv:2407.01502.
- Kolbjørnsrud, V. (2023), 'Designing the intelligent organization: Six principles for human-ai collaboration', *California Management Review* **66**(2), 44–64. Original work published 2024.
URL: <https://doi.org/10.1177/00081256231211020>
- Kozachek, D. (2026), 'Collaborative Agentic Information Systems (CAIS) MCP'.
URL: <https://github.com/koizachek/cais-mcp>
- Kumar, N., Wei, X. & Zhang, H. (2026), 'Agentic artificial intelligence as a new frontier in information systems: Promise, peril, and research opportunities', *Information Management* **63**(3), 104317.
URL: <https://www.sciencedirect.com/science/article/pii/S0378720626000170>
- Leonardi, P. (2023), 'Guest editorial: Affordances and agency: A clarification and integration of fractured concepts', *MIS Quarterly* **47**(4).
- Lou, B., Lu, T., Raghu, T. S. & Zhang, Y. (2026), 'Visioning human-agentic ai teaming: Continuity, tension, and future research'.
URL: <https://arxiv.org/abs/2603.04746>
- Masters, C., Vellanki, A., Shangguan, J., Kultys, B., Gilmore, J., Moore, A. & Albrecht, S. (2025), Orchestrating human-ai teams: The manager agent as aunifying research challenge, in 'Proceedings of the 2025 7th International Conference on Distributed Artificial Intelligence', DAI '25, Association for Computing Machinery, New York, NY, USA, pp. 91–107.
URL: <https://doi.org/10.1145/3772429.3772439>
- Mayer, V., Schüll, M., Aktürk, O. & Guggenberger, T. (2024), Designing human-ai hybrids: Challenges and good practices from a multiple case study, in 'Proceedings of the International Conference on Information Systems (ICIS)', p. 3. ICIS 2024.
URL: <https://aisel.aisnet.org/icis2024/aiinbus/aiinbus/3>
- McHugh, M. L. (2012), 'Interrater reliability: the kappa statistic', *Biochemia Medica* **22**(3), 276–282.
- McNeese, N. J., Schelble, B. G., Canonico, L. B. & Demir, M. (2021), 'Who/what is my teammate? team composition considerations in human–ai teaming', *IEEE Transactions on Human-Machine Systems* **51**(4), 288–299.
- Naik, S., Toombs, A. L., Snellinger, A., Saponas, S. & Hall, A. K. (2025), 'Exploring human-ai collaboration using mental models of early adopters of multi-agent generative ai tools'.
URL: <https://arxiv.org/abs/2510.06224>
- Panzarasa, P., Jennings, N. R. & Norman, T. J. (2002), 'Formalizing collaborative decision-making and practical reasoning in multi-agent systems', *Journal of Logic and Computation* **12**(1), 55–117.
URL: <https://doi.org/10.1093/logcom/12.1.55>
- Rai, A., Constantinides, P. & Sarker, S. (2019), 'Next generation digital platforms: Toward human-ai hybrids', *MIS Quarterly* **43**, iii–ix.
- Rinta-Kahila, T., Penttinen, E., Salovaara, A., Soliman, W. & Ruissalo, J. (2023), 'The vicious circles of skill erosion: A case study of cognitive automation', *Journal of the Association for Information Systems* **24**(5).

- Russell, S. & Norvig, P. (2009), *Artificial Intelligence: A Modern Approach*, 3rd edn, Prentice Hall Press, USA.
- Schmitt, A. (2024), Ensuring human agency: a design pathway to human-ai interaction, in 'ICIS 2024 Proceedings'.
- Schmutz, J. B., Outland, N., Kerstan, S., Georganta, E. & Ulfert, A.-S. (2024), 'Ai-teaming: Redefining collaboration in the digital era', *Current Opinion in Psychology* **58**, 101837.
URL: <https://www.sciencedirect.com/science/article/pii/S2352250X24000502>
- Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G. & Söllner, M. (2020), 'Machines as teammates: A research agenda on ai in team collaboration', *Information Management* **57**(2), 103174.
URL: <https://www.sciencedirect.com/science/article/pii/S0378720619303337>
- Shavit, Y., Agarwal, S., Brundage, M., Adler, S., O'Keefe, C., Campbell, R., Lee, T., Mishkin, P., Eloundou, T., Hickey, A., Slama, K., Ahmad, L., McMillan, P., Beutel, A., Passos, A. & Robinson, D. G. (2023), Practices for governing agentic ai systems, White paper, OpenAI.
URL: <https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf>
- Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G. & Moy, L. (2023), 'Chatgpt and other large language models are double-edged swords', *Radiology* **307**(2), e230163.
URL: <https://doi.org/10.1148/radiol.230163>
- Song, T., Tan, Y., Zhu, Z., Feng, Y. & Lee, Y.-C. (2025), 'Multi-agents are social groups: Investigating social influence of multiple agents in human-agent interactions', *Proc. ACM Hum.-Comput. Interact.* **9**(7).
URL: <https://doi.org/10.1145/3757633>
- Strunk, J., Banh, L., Nissen, A., Strobel, G. & Smolnik, S. (2024), To delegate or not to delegate? factors influencing human-agentic is interaction, in 'ICIS 2024 Proceedings'.
- Sudeeptha, I., Mueller, W., Leyer, M., Richter, A. & Nolte, F. (2025), Use cases for prospective sensemaking of human-ai collaboration, in 'Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS 2025)'.
- Szeider, S. (2025), Bridging language models and symbolic solvers via the model context protocol, in '28th International Conference on Theory and Applications of Satisfiability Testing (SAT 2025)', Vol. 341 of *Leibniz International Proceedings in Informatics (LIPIcs)*.
- Tolzin, A. & Janson, A. (2026), 'Uncovering the mechanisms of common ground in human-agent interaction: review and future directions for conversational agent research', *Internet Research* **36**(1), 292–315.
URL: <https://doi.org/10.1108/INTR-06-2023-0514>
- Tolzin, A., Körner, A., Dickhaut, E., Janson, A., Rummer, R. & Leimeister, J. M. (2023), Designing pedagogical conversational agents for achieving common ground, in 'International Conference on Design Science Research in Information Systems and Technology (DESRIST)', Pretoria, South Africa.
- Viswanathan, P. S. (2025), 'Agentic ai: A comprehensive framework for autonomous decision-making systems in artificial intelligence', *International Journal of Computer Engineering and Technology* **16**(1), 862–880.

- Vössing, M., Kühl, N., Lind, M. & Satzger, G. (2022), 'Designing transparency for effective human-ai collaboration', *Information Systems Frontiers* **24**, 877–895.
- Walls, J. G., Widmeyer, G. R. & El Sawy, O. A. (1992), 'Building an information system design theory for vigilant EIS', *Information Systems Research* **3**(1), 36–59.
- Webster, J. & Watson, R. T. (2002), 'Analyzing the past to prepare for the future: Writing a literature review', *MIS Quarterly* **26**(2), xiii–xxiii.
- Wu, S., Liu, Y., Ruan, M., Vasukin, A., Johansson, S., Bonnin, V., Natarajan, P. & Hoffman, L. (2025), 'Human-generative ai collaboration enhances task performance but undermines human's intrinsic motivation', *Scientific Reports* **15**.
URL: <https://doi.org/10.1038/s41598-025-98385-2>
- Xu, J. J. & Topi, H. (2017), 'A conceptual model for user-system collaboration: Enhancing usability of complex information systems', *Communications of the Association for Information Systems* **41**, 775–800.
URL: <https://doi.org/10.17705/1CAIS.04131>