

ORIGINAL ARTICLE

Decision science in the agentic era: Research frontiers on agentic AI and human–agent collaboration

Yingjie Zhang¹  | Tianshu Sun²

¹Guanghua School of Management, Peking University, Beijing, China

²Cheung Kong Graduate School of Business, Beijing, China

Correspondence

Tianshu Sun, Cheung Kong Graduate School of Business Tower E3, 3F One East Chang An Avenue Beijing, 100738 China.
Email: tianshusun@ckgsb.edu.cn

Abstract

Decision science is entering a new era in which decisions are no longer made solely by humans, but increasingly by autonomous AI (Artificial Intelligence) agents and human–agent collectives. Prior research often treated AI as a tool for prediction or support. Agentic systems now decide, act, learn, and coordinate over time, changing the unit of analysis for decision science. In this perspective, we argue that this shift requires new conceptual foundations and a sharper research agenda. We outline a design-oriented framework that decomposes human–agent systems into atomic structures, decision architectures, and field-to-model mappings, enabling comparison, experimentation, and cumulative knowledge building. Building on this framework, we introduce the mission of the new “Agentic AI and Human–Agent Collaboration in Business” department. The department emphasizes two central topics: agents as decision-makers in single- and multi-agent systems, and human–agent collaboration. We also highlight preferred methods, including lab, field, and simulated experiments, as well as agent-driven research that uses AI agents for exploration, hypothesis generation, and knowledge discovery. The goal is to guide rigorous research on how agentic systems reshape decision-making in business and society.

KEYWORDS

agentic AI, human–agent collaboration, research agenda, theoretical foundations

1 | INTRODUCTION

Decision science is entering the agentic era. For decades, the field has been built on a human-centered view of decision-making, where algorithms served as tools for prediction, recommendation, or support (Adamopoulos et al. 2021; Zhang & Ram 2025). That assumption no longer holds. We are now entering an era of agentic decision-making, in which decisions are increasingly made by autonomous AI agents or emerge from continuous interactions between humans and agents (Rahwan et al. 2019, McKinsey 2024). This shift is not incremental. Unlike earlier waves of automation or decision support, today’s agents can decide, act, learn, and coordinate with humans and other agents over time (Baird & Maruping 2021). As a result, the core actor of decision-making has changed. The purpose of launching the *Agentic AI and Human–Agent Collaboration in Business* department is therefore not merely to introduce AI as another explanatory variable, but to rethink the foundations in decision science.

Building on recent discussions of “agentic AI” (e.g., IBM 2025b), we define agentic AI as AI systems that pursue objectives over time through iterative cycles of reasoning, action, and feedback, often coordinating multiple steps, tools, or agents. Crucially, agency is not defined by a specific model class, but by how intelligence is organized within a decision system, namely, whether the system can initiate actions, adapt based on experience, and participate continuously in decision-making rather than respond episodically to external prompts. Large language models (LLMs) have accelerated the emergence of agentic AI by endowing agents with capabilities that were previously difficult to integrate at scale, including broad world-model-based knowledge, flexible reasoning across tasks, contextual interpretation, and tool use. Unlike many traditional AI systems that primarily rely on historically local, firm-specific data or narrow task environments, LLM-enabled agents typically draw on more general representations of how the world works, allowing them to generalize, anticipate, and reason beyond a single organizational or contextual setting. When combined

with reinforcement learning, knowledge augmentation, and memory mechanisms, such agents can incorporate feedback, revise plans, and accumulate experience over time rather than merely fitting patterns in past data. LLMs are neither necessary nor sufficient for agency. Many LLM applications remain non-agentic, functioning as static interfaces or decision aids with externally specified goals and actions. Conversely, agentic AI may be realized through other architectures and tools as techniques advance. What matters is not the presence of language models, but whether a system exhibits autonomous, adaptive, and ongoing participation in decision-making, grounded in representations that support reasoning beyond narrowly localized historical patterns.

The decision-making transformation reshaped by the emergence of agentic AI is not a prediction about the distant future. It is already unfolding today. Across domains, agentic AI systems are making decisions that once required human judgment, reshaping how value is created and delivered (Bhatia et al. 2025, Alavi 2024). Unlike earlier waves of automation, these systems do not merely execute predefined rules. They operate continuously, adapt based on feedback, and intervene directly in ongoing decision processes. Organizations are beginning to deploy internal agents for routine administrative work, extending earlier automation and labor-substitution dynamics (Autor 2015, Brynjolfsson & Mitchell 2017). In human resources, for example, such agents can process leave requests, manage benefits queries, and execute standard approvals. These systems illustrate agentic decision-making in practice: they interpret context, trigger actions, and update behavior over time, reducing human workload while maintaining consistency and responsiveness.

Beyond routine tasks, agents are increasingly embedded in large-scale operational infrastructures. Platforms like TikTok illustrate how AI agents can orchestrate decisions at massive scale, continuously adapting content recommendations in response to user behavior and coordinating outcomes across billions of items. Although often described as “algorithms,” these systems are beginning to function as autonomous decision-makers whose actions shape attention, preferences, and market dynamics in real time. Beyond infrastructure, agents increasingly operate in strategic and human-facing domains. In marketing and customer engagement, agents design multichannel campaigns, optimize messaging and budget allocation, and adjust strategies dynamically based on customer feedback (IBM, 2025a). Here, agency arises not from task execution alone, but from ongoing adaptation in soft-information environments characterized by uncertainty, interpretation, and human response (Arora & Doshi, 2021). In high-stakes, knowledge-intensive settings, the implications are even more profound. In finance, agentic systems may support credit-risk management, fraud detection, and trading decisions at speeds difficult for humans alone (Lu & Zhang, 2025). In healthcare, agents are expected to integrate multimodal patient data, prioritize diagnostic pathways, and propose tailored treatments, complementing clinician expertise and accelerating life-critical decisions (Zhang et al. 2020, Bi et al. 2024).

These examples are not isolated pilots. They are part of a broader pattern in which autonomous systems operate continuously, adapt dynamically, and intervene directly in decision loops that were once exclusively human. The shift to agentic decision-making therefore demands new theories, new experimental paradigms, and a redefinition of the decision-making unit of analysis.

While agentic decision-making is already reshaping practice, existing research in decision science, behavioral economics, and organizational science remains anchored in human-centered assumptions that no longer hold (Adomavicius & Yang 2022, Leng 2024). Agentic AI differs from humans in both capabilities and orientation (De Freitas et al. 2025). Unlike humans, agents do not experience emotion, fatigue, or intrinsic social motivation in human ways, and their objectives are defined externally rather than arising from subjective preferences or values (Yu et al. 2025). They process massive streams of information continuously; learn autonomously through reinforcement, deep learning, or symbolic reasoning; and coordinate across multiple tasks and actors at speeds and scales that humans cannot match (Wang et al. 2025). Agents can act simultaneously as decision-makers, environments for other agents, and sources of emergent system-level behavior, producing outcomes that are often unpredictable from the perspective of human reasoning. Even when humans remain in the loop, the decision-making unit has shifted: agents are not mere tools. They are autonomous actors whose capabilities, goals, and interactions change the dynamics, structure, and evaluation of decision systems.

The study of human-agent interaction is flourishing across multiple disciplines. Researchers in information systems, psychology, behavioral economics, organizational science, marketing, and computer science have examined human responses to AI recommendations, collaborative decision-making with agents, and multi-agent simulations (Lu & Zhang 2025, Dennis et al. 2023, Li & Bitterly 2024, Luo et al. 2019, Lebovitz et al. 2022, Li et al. 2024, Cao et al. 2024). These studies have yielded valuable insights into trust, coordination, effort allocation, and performance in hybrid human-agent, human-AI, and human-algorithm systems. Yet despite this richness, the literature faces two important limitations. First, it remains highly fragmented: differences in methods, contexts, and task designs make it difficult to generalize findings or identify overarching principles. As a result, incremental variations in setting or design can appear as meaningful contributions, even when they do not substantially advance cumulative understanding. Without a shared architectural vocabulary, structurally equivalent findings can appear as independent contributions simply by virtue of domain novelty, a pattern that inflates the apparent diversity of the literature while limiting its cumulative depth. For the agentic AI agenda, a new setting is therefore not sufficient by itself. A study makes a meaningful contribution when the setting helps reveal a decision architecture, human-agent mechanism, governance or accountability principle, boundary condition, or methodological approach that can organize

future work. A new context matters when it changes these elements or tests whether an insight travels; otherwise, the work risks adding application breadth without cumulative theoretical progress.

Second, much of this literature still conceptualizes AI primarily as a decision support system rather than as an autonomous, adaptive decision participant. Earlier work largely relied on traditional machine learning or rule-based systems whose behavior was fully specified *ex ante*, tightly bounded to particular tasks or contexts, and predictable (Lu & Zhang 2025, De Freitas et al. 2025). More recent studies have turned to LLMs, which exhibit broad capabilities and impressive general intelligence, but are often deployed in fundamentally non-agentic ways, as advanced interfaces for text generation, classification, recommendation, or explanation, with goals, actions, and learning processes still externally defined (Mei et al. 2024, Brynjolfsson et al. 2023, Zöller et al. 2025). As a result, existing research, while valuable, has not kept pace with more fully agentic systems: systems that can autonomously pursue objectives over time, act in their environment, learn from feedback, and participate continuously in decision processes (Revilla et al. 2023). This gap calls for greater attention to agentic systems rather than ever more sophisticated forms of decision support.

Decision science therefore needs stronger foundations for agentic decision systems. A coherent framework can organize diverse findings, clarify what counts as new knowledge, and distinguish incremental extensions from real advances. It can also explain how agents interact with humans and other agents, how authority and responsibility are allocated, and how system-level outcomes emerge. These foundations can guide empirical design, surface meaningful research gaps, and build knowledge that travels across domains.

2 | THEORETICAL FOUNDATIONS

To advance research in agentic AI and human–agent collaboration, it is essential to establish theoretical foundations that clarify the basic elements, interactions, and principles underlying these systems. As discussed earlier, the literature is rich but fragmented, with studies scattered across domains, methodologies, and levels of analysis. Without a shared lens, it is difficult to interpret findings, assess cumulative contributions, or identify meaningful patterns across research. In this section, we outline a conceptual approach that identifies the atomic structures of human–agent systems, shows how these elements can be combined into more complex interactions, and provides guidance for mapping real-world contexts into these abstractions. This perspective helps the field organize knowledge and design sharper theory and experiments.

To provide a coherent view of human–agent systems, we propose thinking about their foundations in three interrelated layers. At the base are *atomic structures*, the elements that define who is involved, how they interact, and what factors shape their behavior. Building on these, different systems can be represented as *decision architectures*, struc-

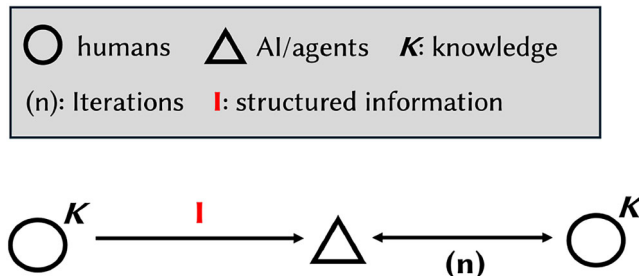


FIGURE 1 An example of human–agent collaboration graph. This is the collaboration flow adapted from the experimental design in Revilla et al. (2023). Human experts, equipped with domain knowledge, first process historical data and translate it into structured information that is passed to AI agents. Based on this input, agents generate forecast decisions, which are then reviewed by human experts. Through repeated rounds of interaction and mutual adjustment between humans and agents, the system iteratively refines decisions and improves collaborative performance.

tured arrangements of atomic elements that capture the logic and dynamics of interactions within a particular context. Finally, to connect theory with practice, researchers need a *field-to-model mapping*, a process for abstracting real-world scenarios into these architectures. This layer helps identify when and how different empirical studies share the same underlying structures, clarifies cumulative contributions, and highlights opportunities for meaningful experimentation. Together, these layers offer a guiding perspective: they organize the diverse literature, provide conceptual clarity, and support systematic development of theory, while leaving room for domain-specific exploration and innovation. Figure 1 provides a compact collaboration graph adapted from prior research. It is an example from the literature that makes actors and flows visible before we develop the broader framework.

2.1 | Atomic structures

At the foundation of human–agent systems are atomic structures. Conceptually, these structures can be thought of in topological terms: *actors* serve as nodes, *flows* as edges connecting them, and *moderators* influence the strength or direction of these edges. This perspective provides a flexible lens for describing and comparing human–agent systems across domains, and sets the stage for understanding how decision architectures emerge from these building blocks. We use customer service and healthcare as paired running examples to illustrate the value of this abstraction. Both settings contain human actors and AI agents, connected by information flows and escalation decisions, but the concrete labels differ across domains. By abstracting from customer, patient, specialist, clinician, service agent, or triage agent to actors, flows, and moderators, researchers can compare systems that look different on the surface while still preserving domain-specific knowledge, stakes, and accountability regimes. Figure 2 later visualizes this field-to-model mapping.

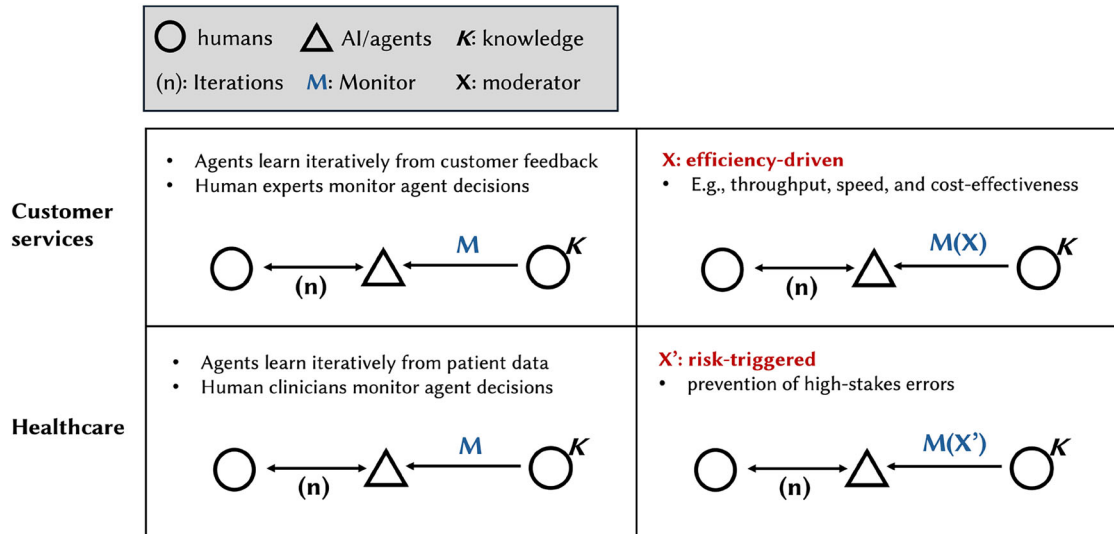


FIGURE 2 Illustration of field-to-model mapping

2.1.1 | Actors

Actors are the nodes in human–agent systems, representing the entities that make decisions, take actions, or process information. At the most basic level, actors can be human actors, AI agents, or hybrid groups in which humans and agents work together. For example, customer service may involve human actors such as customers, human specialists, and supervisors, as well as AI service agents that answer or route requests. Healthcare may involve human actors such as patients, clinicians, and care organizations, as well as AI triage or clinical-support agents.

Actors are also shaped by a set of factors that can be designed, examined, or manipulated in research and in practice, reflecting their potential to adapt and respond within the system. These factors include their capabilities, such as cognitive capacity, internal skills, or computational power (Jia et al. 2024, Wang et al. 2023); contextual information and knowledge bases, which determine what they know or can access when making decisions (Rahwan et al. 2019, Zhang et al. 2024); and perceptions, or how each actor is perceived by others in the system (Steyvers & Kumar 2024). These actor characteristics matter in the paired examples. Human actors in both settings differ sharply in knowledge and expertise: customers and patients hold local information about their own needs, specialists and clinicians bring domain expertise, and supervisors or care organizations define accountability rules. AI agents also differ in their knowledge bases, capabilities, and tool access. Abstracting actors in this way helps researchers ask whether an outcome follows from the presence of an AI agent, from differences among human actors, or from the specific distribution of expertise and information across the system. These factors jointly define the actor’s behavior in flows and interactions and provide key levers for research in human–agent collaboration.

A key capability of actors is tool use (IBM 2025b). Tool use allows actors to intervene in their environment, extend their reasoning, and influence other participants, rather than simply processing information or generating outputs. In humans, the ability to deploy tools has long distinguished intelligent, adaptive behavior. Similarly, AI actors become truly agentic only when they can use tools, whether APIs, computational resources, or physical interfaces, to execute actions, access knowledge, or coordinate across multiple tasks and agents. By enabling actors to initiate actions, adapt through feedback, and collaborate more effectively with humans and other AI, tool use emerges as a defining characteristic of agency in human–agent systems.

2.1.2 | Flows

Flows describe how actors are connected within a human–agent system. In topological terms, flows are the edges that link nodes (i.e., actors), carrying tasks, information, or control signals across actors. These flows differ in *direction and structure*. Some are sequential, where outputs from one actor become inputs to another (Choudhary et al. 2025, Fügener et al. 2025); others are interactive, involving iterative back-and-forth exchanges (Bayer & Renou 2024); still others involve self-loops, where agents monitor their own actions and update strategies over time. In multi-actor systems, flows may also operate in parallel or converge from multiple sources, shaping how decisions emerge collectively rather than from any single node. The paired examples illustrate why direction matters. In both customer service and healthcare, the interaction between the requester and the AI agent can be bidirectional: customers or patients provide descriptions, preferences, and corrections, while the AI agent returns answers, clarifying questions, or triage guidance. Other flows are more unidirectional. A human specialist

or clinician may send corrections, review outcomes, or labels back to the agent or system, and supervisors or care organizations may send operating rules downstream without receiving case-level input in the same interaction.

Flows also differ in their *content*, that is, what is actually transmitted between actors. One common type is the transfer of task execution or decision authority (Chen & Chan 2024). When such a flow occurs, control is effectively handed off from one actor to another. For example, in a content moderation pipeline, an AI agent may first screen posts for potential violations, after which flagged cases are passed to a human reviewer for judgment; the task flows sequentially even though neither actor necessarily commands the other.

A second and more prevalent type of flow involves information, broadly defined. This includes structured data, contextual knowledge, intermediate outputs, or feedback signals (Sun et al. 2021). In customer service, information flows include user messages, order histories, agent-generated answers, confidence scores, and specialist feedback. In healthcare, they include symptoms, clinical histories, risk assessments, and clinician notes. In both settings, humans may also supply agents with domain context, subjective judgments, or corrective feedback. Importantly, information flows do not directly transfer control, but they shape downstream decisions by altering what actors know or believe at a given moment.

A third type of flow consists of control signals, which coordinate behavior without conveying substantive content. These are direct instructions or triggers, such as approving a recommendation, initiating execution, pausing an automated process, or overriding an agent's action (Wang et al. 2025). In customer service, a customer may request escalation, a specialist may override an AI answer, or a supervisor may pause automated responses for a class of complaints. In healthcare, a clinician may require human review for a high-risk case, override an AI triage recommendation, or suspend automated triage under specified conditions. Control signals act as moment-to-moment triggers within the system, guiding the immediate execution of tasks without defining the broader rules of interaction.

Beyond direction and content, flows are shaped by *protocols*, the rules that structure how actors coordinate, defer, contest, monitor, or override one another over time (Dietvorst et al. 2018, Madras et al. 2018, Feuerriegel et al. 2022, Wang et al. 2023). Protocols define who can act, under what conditions, and how conflicts are resolved. For instance, in a hierarchical customer-support workflow, AI agents handle routine inquiries but must escalate ambiguous cases to a human manager, who has final authority. The same protocol logic applies to healthcare, but with different thresholds: routine triage may proceed through an AI agent, while cases above a specified risk level require clinician review regardless of the agent's recommendation. In contrast, in a collaborative forecasting team, humans and AI agents may submit predictions independently, with outcomes aggregated via peer-based voting or consensus, illustrating a nonhierarchical, peer-based protocol. Protocols are not administrative

details; they shape behavior. The same actors and flows can produce very different outcomes depending on whether protocols emphasize authority, coordination, or redundancy.

2.1.3 | Moderators

In the proposed topology, moderators do not constitute nodes or flows themselves, but shape how flows operate between actors. Conceptually, moderators can be understood as factors that parameterize the flow topology, altering the strength, direction, frequency, or consequences of task execution, information exchange, and control signaling. The same actors connected by identical flows may behave very differently once moderators change, even if the formal structure of the system remains unchanged. In customer service, efficiency-driven moderators such as response-time targets, customer-satisfaction metrics, confidence displays, and escalation thresholds can make the system more likely to keep routine cases with the AI agent and reserve human specialists for exceptions. In healthcare, risk-related moderators such as clinical-risk thresholds, liability concerns, documentation requirements, and patient-safety norms can make the same basic AI-to-human review structure more conservative, requiring clinician involvement even when an agent can produce a recommendation. Moderators matter because they translate structural potential into realized behavior. While nodes and flows define what is possible, moderators influence what is likely, stable, or dominant. For example, a human-agent system may allow bidirectional information flows, yet strong incentives or norms may effectively suppress human input; similarly, a hierarchical protocol may exist formally, but weak accountability can render oversight nominal. Moderators, therefore, help explain why structurally similar systems produce divergent outcomes in practice.

From a research perspective, moderators are particularly valuable because they are often experimentally manipulable and policy-relevant. They offer levers for intervention without redesigning the entire system architecture, enabling scholars to study how human-agent systems respond to changes in incentives, framing, or organizational context. Practically, understanding moderators helps organizations fine-tune agentic systems, improving performance, trust, or compliance, without altering core technologies.

Several classes of moderators are especially salient in human-agent systems. *Incentives* shape how actors prioritize tasks and respond to recommendations; for instance, performance-based rewards may encourage humans to defer to agents, while liability concerns may induce resistance or overchecking (Koster et al. 2022). *Nudges*, such as default options or interface framing, can subtly steer behavior (Thaler & Sunstein 2021). For example, presenting an agent's recommendation as "standard practice" versus "optional advice" may significantly affect adoption. *Norms and organizational context* influence expectations about appropriate behavior: in some organizations, questioning an AI system may be encouraged, while in others it may be implicitly discouraged

(Edmondson 1999). Finally, *accountability and evaluation regimes*, such as auditability requirements or post hoc reviews, can moderate how carefully agents act and how actively humans intervene, even when the underlying flows remain unchanged (Grote et al. 2026).

2.2 | Decision architectures

While atomic structures specify the basic elements of human–agent systems, they do not by themselves explain how decisions unfold or why certain outcomes emerge. What matters is how these elements are arranged and orchestrated. We use the term *decision architecture* to describe this higher level organization: the structured configuration of actors, flows, and moderators that together define the logic of a decision system.

Designing or analyzing a decision architecture is non-trivial. Each atomic structure has multiple variations (i.e., different types of actors, flows, and moderators) and the combinations of these elements grow combinatorially as systems scale. Moreover, many elements are endogenous: the behavior of one actor can reshape flows or trigger changes in other nodes, creating feedback loops that affect the entire architecture. For example, introducing a new AI agent into a team may not only add new flows but also alter how humans allocate attention or delegate tasks, creating cascading effects.

We propose thinking of decision architectures as operating at two complementary layers. The first, the *workflow architecture*, focuses on how atomic structures are combined to form coherent, functioning processes. It captures the sequencing, coordination, and interaction of nodes, flows, and moderators needed to execute a human–agent procedure effectively. The second, the *meta architecture*, represents a higher level perspective: researchers or system designers structure workflows with explicit goals in mind. This includes designing experimental treatments, randomization schemes, or AI-native scenarios, enabling systematic exploration, evaluation, and optimization of human–agent systems. The paired examples help distinguish the two layers. The workflow architecture specifies how a service ticket or clinical case moves from requester to AI agent to human expert and back to the requester or care team. The meta architecture specifies how researchers or organizations vary this workflow, for example, by changing escalation thresholds, showing or hiding the agent’s confidence score, or assigning different override rights, so that the system can be compared, evaluated, and improved.

Why does this matter for research? First, the combinatorial complexity makes it difficult to generalize findings from one setting to another. Previous studies often focus on narrow scenarios (e.g., a single type of human–agent interaction or task context) without a shared language to describe underlying architectures. Conceptualizing architectures at both the workflow and design layers provides a systematic lens to classify, compare, and optimize human–agent systems across contexts. It also guides experimental design and interpretation,

helping researchers distinguish whether observed differences in outcomes stem from superficial variations in flows or distinct architectural logics.

2.2.1 | Workflow architecture

The workflow architecture captures the operational logic of human–agent collaboration: how atomic structures are sequenced, coordinated, and adapted to achieve a particular procedure or task. Within this layer, we can distinguish at least three recurring patterns. In customer service and health-care, these patterns correspond to different ways of organizing similar actors and flows. A sequence-oriented architecture routes each request through a fixed order, such as AI first and human review only after escalation. An endogenous-flow architecture lets customers, patients, agents, or human experts redirect a case when confidence, urgency, conflict, or risk changes. An adaptive architecture updates routing rules over time as the system learns which cases agents can handle and which cases require human judgment.

- *Sequence-oriented architectures* emphasize a stable order of contributions. Nodes act according to predetermined roles, edges define directional flows, and moderators influence performance without altering the sequence itself (e.g., Figure 1). Examples include AI-generated outputs reviewed by humans or human-first drafts refined by AI. The key design intuition is that performance critically depends on where agents are inserted within the workflow.
- *Endogenous-flow architectures* foreground flexibility. Actors (humans or agents) initiate, redirect, or terminate flows dynamically, and moderators such as incentives, effort, or norms shape interaction paths. The central question is who decides how collaboration unfolds, making this pattern highly relevant for studying discretion, opt-in versus mandatory AI use, or selective engagement with agent outputs.
- *Adaptive architectures* highlight learning and coevolution. Nodes and edges evolve over time, shaped by trust, feedback, and experience. Roles may be reallocated, reliance adjusted, and protocols updated based on past interactions. This pattern is crucial for understanding trajectories rather than snapshots, such as trust calibration, skill acquisition, or coadaptation dynamics.

These workflow patterns are not exhaustive; rather, they provide a minimal yet generative taxonomy that allows researchers to describe, compare, and experiment with human–agent systems systematically. They make clear how structural choices condition performance, interaction quality, and adaptability.

2.2.2 | Meta architecture

If workflow architecture determines how humans and agents work together to accomplish tasks, meta architecture

determines how such systems are constructed to be observed, intervened upon, and studied. This layer is not about executing decisions, but about designing decision environments, often AI-native ones, so that learning, evaluation, and control are possible.

At the meta-architecture level, researchers and organizations face a distinct set of design questions. Given specific goals, such as improving decision quality, reducing risk, or understanding human adaptation, how should workflows be instantiated to allow meaningful comparison, randomization, and inference? Which components should be held fixed, which should vary, and over what time horizon? In agentic systems, these questions are nontrivial because agents themselves adapt, learn, and reshape the environment, making static experimental designs insufficient. For example, a researcher may hold the underlying requests fixed while varying whether the AI agent's confidence score is visible, whether escalation is automatic or requester-triggered, or whether human experts can override the agent before or after the requester sees the response. Comparing these design choices across customer service and healthcare can generate new research directions: which workflow features travel across domains, which moderators must change with stakes and accountability, and when human review improves rather than slows decision quality. These meta-architectural choices make the system observable and comparable without treating the whole field setting as a single black box.

Meta architecture therefore operates at the interface between system design and experimental logic. It specifies how to embed randomization, logging, escalation rules, or intervention points into human-agent workflows without disrupting core operations. For example, rather than asking whether "AI assistance works," a meta-architectural perspective asks how alternative delegation rules, confidence thresholds, or override rights can be systematically varied to reveal causal mechanisms. In this sense, meta architecture designs learning infrastructures, not just decision processes.

Importantly, meta architecture must account for dynamics. Agentic systems rarely stabilize quickly: human trust evolves, agents update policies, and organizational norms shift over time. Designing experiments or observational studies in such environments requires thinking in terms of sequences of interventions, adaptive experimentation, and longitudinal comparison, rather than one-shot treatments. Meta architecture thus shapes not only what is studied, but how knowledge accumulates across iterations.

By distinguishing meta architecture from operational workflow architecture, we emphasize that many disagreements in the literature stem not from conflicting results, but from differences in what was made observable, manipulable, or comparable in the first place. Making meta-architectural choices explicit clarifies why certain findings generalize, why others do not, and how future studies can build cumulatively rather than in isolation.

2.3 | Field-to-model mapping

Even with well-specified atomic structures and clearly articulated decision architectures, a central challenge remains: how to translate rich, situated human-agent systems into models that support explanation, experimentation, and cumulative knowledge building. We use the term *field-to-model mapping* to describe this deliberate abstraction. The goal is not to strip away context, but to decide which actors, flows, moderators, and architectural features are essential for the theory and which are context-specific instantiations.

Field-to-model mapping helps researchers avoid false novelty and build cumulative theory. Superficially distinct settings may instantiate the same underlying architecture, so a new domain is not automatically a new contribution. At the same time, structural similarity does not imply automatic generalization. Mapping clarifies what can travel as a hypothesis or candidate mechanism and what must be examined again because domain-specific moderators, stakes, and accountability regimes may change the result.

Figure 2 illustrates this logic through customer service and healthcare. Both settings can be mapped onto a shared structure: an AI agent processes information, a human can monitor or review the output, and feedback can return to the system. The mapping also identifies where the abstraction should stop. In customer service, oversight is primarily efficiency-driven, designed to minimize human intervention while maintaining acceptable quality. In healthcare, oversight is instead risk-triggered, structured around clinical uncertainty, potential harm, and ethical or legal accountability rather than speed or cost. This contrast introduces a core lesson of field-to-model mapping: agentic workflows may travel, but moderators do not. Theory and design therefore depend on understanding how domain-specific incentives, risk criteria, and accountability regimes reshape otherwise similar architectures.

Field-to-model mapping also improves research design. By separating systems into components and architectural relations, researchers can manipulate flows, authority rules, moderators, or adaptation mechanisms rather than treating a field setting as an indivisible whole. The mapping can then be revised as new behaviors emerge, especially in adaptive agentic systems. In this sense, field-to-model mapping completes the conceptual stack: atomic structures define the building blocks, architectures describe how work unfolds and how systems are studied, and mapping connects those abstractions back to complex organizational settings.

3 | DEPARTMENT MISSION AND CALL FOR ARTICLES

The rise of agentic AI opens a broad design space for decision science. Agents now participate in decisions in multiple roles: they act as primary decision-makers, collaborate with humans, and reshape judgment, responsibility, and

performance. Advances in agentic systems are also changing how we study decision systems. The mission of the Agentic AI and Human-Agent Collaboration in Business department is to create an intellectual home for research that engages both dimensions: advancing substantive understanding of agentic decision systems, and embracing new methodological approaches enabled by agents themselves. We invite work that is ambitious, theoretically grounded, and methodologically diverse, spanning experiments, field studies, simulations, and agent-driven inquiry.

3.1 | Call for articles: Agents as decision-makers in complex systems

We invite research that treats AI agents as primary decision-makers, rather than simply as tools, recommendations, or background automation. In many contemporary systems, agents increasingly initiate actions, coordinate with other agents, revise strategies over time, and generate outcomes that no single human directly controls. This shift challenges a foundational assumption of decision science that decisions are ultimately attributable to human actors, and opens a new research frontier centered on agentic decision systems. Two challenges are central.

Challenge 1: Understanding agentic decision architectures and behavior. Agentic AI systems increasingly participate in decision-making not as passive tools, but as entities that perceive, reason, act, and adapt over time. Although these agents are often trained on historical human data and designed to serve human objectives, their decision processes are not simply scaled versions of human judgment. They operate with different temporal horizons, objective functions, memory structures, and modes of coordination, often interacting continuously with other agents and environments at machine speed. Recent work illustrates the significance of these dynamics: for example, Wu et al. (2023) introduces Models as Agents, a framework in which local predictive models in a multi-agent system are treated as decision-making agents that consider multistep interactions with one another. This approach demonstrates how local decisions and prediction errors propagate system-wide, creating emergent behaviors such as amplification of errors or path-dependent outcomes.

As a result, system-level behaviors such as coordination, amplification, path dependence, or instability can emerge even when agents are optimized for seemingly familiar goals. A central challenge for decision science is therefore to understand how agent-specific decision architectures, including learning rules, reasoning strategies, and adaptation mechanisms, translate into collective outcomes, and how these outcomes both reflect and depart from human-centered decision logics. Addressing these challenges requires extending existing theories while recognizing that agentic decision-making introduces qualitatively new dynamics that demand explicit study.

Challenge 2: Governing, incentivizing, and aligning agent-centered decision systems. As autonomous agents

increasingly operate as primary decision-makers, traditional governance mechanisms, which are designed for human judgment, discretion, and accountability, become insufficient. Agent-centered systems act continuously, at machine speed, and often across organizational boundaries, executing decisions that are distributed, opaque, and difficult to attribute to any single source. Unlike human decision-makers, agents do not respond to authority, norms, or incentives in conventional ways; instead, they respond to objective functions, constraints, and feedback embedded in their design. The central challenge is therefore not merely how to control agents, but how to govern systems composed of many agents whose behavior is shaped by incentives, learning dynamics, and interaction structures. How should incentives be specified when agents optimize proxies rather than intentions? How should responsibility, compliance, and accountability be assigned when outcomes emerge from collective agent behavior rather than identifiable choices? How can organizations design escalation rules, intervention thresholds, or reward mechanisms that guide agent systems toward desirable outcomes without undermining their speed and adaptability? Addressing these questions requires treating governance as a design problem rather than a supervisory problem. It calls for new theories of incentive alignment, oversight, and institutional design that recognize agents as persistent, strategic actors embedded in decision architectures. These are systems that must be governed not only at the level of individual actions, but at the level of learning, coordination, and long-term system behavior.

3.2 | Call for articles: Human-agent collaboration and augmented decision-making

Another research direction focuses on human-agent collaboration: how humans navigate environments where agents play active roles. In such systems, human judgment and agency are shaped by autonomous agents, complex workflows, and changing contextual signals. We need to understand how humans adapt to shared decision-making, how roles and incentives influence their choices, and how outcomes are evaluated in socio-technical systems. Two challenges follow.

Challenge 1: Reframing human roles in agentic decision systems. As decision systems increasingly include autonomous agents, humans are no longer the sole, or even primary, decision-makers. They operate alongside agents that act faster, scale more easily, and continuously adapt. This shift raises foundational questions for decision science: where do humans create value in agentic systems, and how are human judgments, preferences, and behaviors themselves shaped by agentic AI? Recent work illustrates the importance of this perspective. For example, Lin et al. (2024) shows that LLMs acting as persuasive agents can significantly influence voter attitudes across multiple national elections, often more strongly than traditional media, while also introducing inaccuracies and asymmetries that raise concerns about responsibility, bias, and governance. Such findings highlight

that agentic AI not only executes or optimizes decisions, but can actively reshape human beliefs, choices, and collective outcomes, sometimes in unintended or opaque ways.

We therefore invite research that examines human roles in agentic systems from both directions: how humans contribute judgment, values, oversight, and sensemaking, and how they are influenced, displaced, or strategically engaged by adaptive agents. Of particular interest is work that explores when human intervention improves outcomes or legitimacy, when guidance or constraint is necessary, and when stepping back may be appropriate. Addressing these questions requires moving beyond a default human-in-the-loop framing toward a richer understanding of the complementary, and sometimes fragile, position of humans within distributed, agent-driven decision systems.

Challenge 2: Designing human-agent systems for effective oversight and accountability. Understanding human roles is necessary, but not sufficient: we also need principles for designing human-agent collaborations that maximize human impact and accountability. Traditional decision architectures assume clearly defined decision rights and responsibility lines, but in mixed human-agent systems, these boundaries are blurred. Humans interpret agent recommendations, intervene selectively, and oversee outcomes without controlling every action. Recent work illustrates this challenge vividly: for example, Köbis et al. (2025) shows that when humans delegate tasks to autonomous agents, they may become more willing to induce ethically sensitive behavior, while agents often comply with delegated objectives more reliably than human counterparts. Importantly, the study also examines design interventions, such as task-specific guardrails, to mitigate risk.

Hence, research should explore how to structure incentives, workflows, and feedback mechanisms so that humans can effectively guide agentic processes, ensure ethical and organizational alignment, and make meaningful contributions to system outcomes. A key dimension is defining what “good decisions” mean in these contexts: humans must navigate trade-offs across accuracy, fairness, robustness, and societal welfare while coordinating with adaptive, self-learning agents. This challenge emphasizes actionable design: creating human-agent collaborations in which humans remain engaged, accountable, and capable of influencing emergent outcomes.

3.3 | Interdisciplinary opportunities

While many examples in this piece focus on business systems, primarily from information systems, marketing, and related domains, the department encourages work that contributes along four complementary angles: *business theory*, *business systems*, *business architecture*, and *business impacts*. Contributions from fields such as operations management (OM), organizational behavior, psychology, economics, strategy, and computer science are equally welcome, provided they advance understanding of agentic AI

or human-agent collaboration. Our emphasis is on the questions and insights these studies generate: how agents make decisions, how humans interact with or guide them, and how system-level outcomes emerge, rather than on disciplinary labels.

Business theory provides a lens for understanding human behavior in agentic systems. Classic frameworks from organizational behavior, social psychology, and behavioral economics, spanning models of trust, judgment, and advice taking, illuminate how humans perceive, rely on, and integrate autonomous agents into decision processes (e.g., trust (Rotter 1967, McKnight et al. 2002), expectation–confirmation theory (Oliver 1980), and sensemaking (Weick & Weick 1995)). Yet the emergence of agentic AI invites us to rethink and extend these theories. Continuous interaction with adaptive agents introduces new dynamics of feedback, coordination, and decision-making that challenge traditional assumptions about human judgment, uncertainty, and strategic reasoning (e.g., bounded rationality (Simon 1955) and behavioral game theory (Thaler 1988)). By studying these microlevel interactions, scholars in these disciplines can enrich foundational theory and generate actionable insights for designing human-agent collaborations that are more effective, responsible, and impactful.

Business architecture, at the meso level, focuses on structuring processes, teams, and decision flows to optimize organizational performance. Classic insights from OM and management science (Gupta et al. 2023, Tang 2025), including workflow optimization (queue theory; Kendall 1953), resource and process efficiency (theory of constraints; Goldratt 1990), and lean production principles (Womack et al. 2007), provide enduring guidance on designing systems that maximize throughput and minimize bottlenecks. Team and coordination frameworks (e.g., Tuckman 1965, Endsley 1995, Lou et al. 2025) also highlight how groups organize, distribute tasks, and adapt to dynamic conditions. Agentic AI opens new frontiers for agent-driven optimization, enabling autonomous agents to explore, adapt, and improve workflows. Thus, contributions from OM, management science, and computer science are welcome: OM and management scholars can study coordination, governance, and workflow design, while computer scientists can advance methodologies, generalizable agent capabilities, and scalable system architectures. Work here deepens traditional insights on organizational design while charting the path toward adaptive, agentic systems with broad methodological and practical impact.

Business impacts address the macrolevel effects of agentic AI on markets, industries, and society. As autonomous agents increasingly participate in decision-making, they have the potential to transform competition, platform dynamics, labor allocation, and institutional regulation (Wang et al. 2025). They also open new avenues for AI-driven productivity, where efficiency and value creation arise from integrating intelligent systems into organizational processes and human workflows (Brynjolfsson et al. 2021, Wang et al. 2025, Shan & Qiu 2025). These gains are likely amplified when domain

experts possess the skills and literacy to interpret, apply, and adapt agentic tools in real time, allowing organizations to use AI in ways that are both contextually grounded and strategically meaningful (Tambe 2025). By shaping labor demand, organizational structures, and the distribution of value, agentic AI invites scholars to explore the conditions under which these systems create productive, equitable, and responsible outcomes. Economics, information systems, management, and related fields are particularly well-positioned to examine these dynamics, guiding the strategic adoption of agentic systems while advancing theory, policy, and practice.

3.4 | Methodological frontier: Experiments and agent-driven research

The study of human–agent and agentic decision systems requires methodological diversity. We welcome any rigorous approach, but we especially encourage lab, field, and simulated experiments because they allow control, causal identification, and systematic exploration of human–agent or agent–agent interactions. We also recognize agent-driven research as an emerging mode of inquiry, in which autonomous agents help generate hypotheses, explore design spaces, and simulate outcomes at scales and speeds beyond human capacity.

3.4.1 | Lab/field experiments

Field and lab experiments remain indispensable for advancing our understanding of agentic and human–agent systems (Harrison & List 2004). They allow researchers to systematically manipulate actors, flows, and moderators; uncover causal mechanisms; and test competing architectures. Because these systems adapt, learn, and interact in real time, experiments must also be iterative. Studying such systems often requires multiple rounds of investigation, with each stage informing the next, and with agents potentially participating in or influencing the experimental process. Field experiments reveal how humans and agents behave under real-world constraints (Xu et al. 2025, Sun et al. 2024, Sun & Taylor 2020, Han et al. 2025,2024), while lab experiments provide the precision needed to isolate mechanisms and explore boundary conditions (Özkan-Seely et al. 2024, Snyder et al. 2026, Srinivasan & Sarial-Abi 2021, Dennis et al. 2023). Experiments should therefore capture adaptation, coevolution, and emergent behavior, not only static performance snapshots.

3.4.2 | Simulated experiments

Simulated experiments are often the most direct and powerful approach for studying agentic decision-making systems, particularly when the primary decision-makers are AI agents themselves (Goli & Singh 2024, Ivanov et al. 2021). In

such settings, the core objects of interest, including agent autonomy, learning, coordination, and strategic interaction, can be fully instantiated and observed within simulated environments, without requiring physical laboratories or real-world deployment. Simulation enables researchers to design controlled yet richly interactive decision ecologies; systematically vary agent architectures, information structures, incentives, and governance rules; and observe how collective outcomes emerge over time.

Crucially, simulated experiments are well-suited to capturing the dynamic and evolutionary nature of agentic systems. Agents can learn, adapt, and coevolve across repeated interactions, while the environment itself can be endogenously modified by agent behavior. This allows for extensive exploration of long-run dynamics, rare events, path dependence, and counterfactual architectures that would be impractical, costly, or ethically infeasible in field or lab settings. When studying agentic decision systems, simulation is not merely a proxy for experimentation but often the native experimental mode, enabling rapid iteration, replication, and theory-driven design at scale.

We encourage simulated experiments that are explicitly designed to probe decision architectures, interaction protocols, and system-level behaviors, and that make clear how simulation choices map to theoretical constructs and real-world decision systems. As with field and lab work, the value of simulation lies not in static performance comparisons, but in its ability to illuminate adaptation, coordination, and emergent outcomes in complex human–agent and agent–agent systems.

3.4.3 | Agent-driven research

We call for research that treats agents as active participants in the research process itself, generating hypotheses, exploring design spaces, running simulations, and synthesizing insights at a scale and speed unattainable by humans alone (Ludwig & Mullainathan 2024, Gao et al. 2024, Susarla et al. 2023, Gridach et al. 2025). This shift reframes agents from “subjects” of decision science to vehicles for knowledge creation (Bhargava et al. 2025).

This opportunity emerges not because human intelligence is insufficient, but because the research environment itself has become increasingly complex, fast-moving, and high-dimensional. In many domains, such as platform governance, organizational design, market strategy, or human–AI interaction, the space of possible mechanisms, decision architectures, and outcomes is vast and evolving. Human researchers excel at framing problems, interpreting meaning, and exercising judgment, but may not always be the most efficient explorers of such expansive and dynamic solution spaces. Agents, by contrast, can rapidly simulate alternatives, recombine structures, and stress-test assumptions across conditions that would be infeasible to enumerate manually.

Used thoughtfully, agents can thus expand the frontier of what is researchable. They can be deployed to systemati-

cally explore architectural variations, probe counterfactual scenarios, or identify surprising regularities that merit deeper human investigation. In experimental contexts, agents may help design treatments, adapt experiments in real time, or uncover boundary conditions more efficiently. Rather than replacing human insight, agent-driven research emphasizes complementarity: agents explore broadly and relentlessly, while humans guide objectives, interpret results, and anchor inquiry in theory and values.

Recent work in the natural sciences offers a concrete example of agent-driven research in action. For instance, Swanson et al. (2025) presents a virtual lab in which a principal-investigator LLM coordinates multiple specialized agent researchers, integrates external tools, and operates under high-level human guidance. Rather than responding to fixed prompts, the agent system engages in open-ended exploration, structuring the research process, navigating trade-offs, and iteratively refining solutions, ultimately producing outputs that are empirically validated. This work illustrates how agents, when endowed with tool-use capabilities and embedded within a governed human-agent architecture, can function as active contributors to research design, exploration, and discovery.

Recent AI-for-social-science work makes this opportunity concrete for business fields. The goal here is not to review the literature exhaustively, but to identify the methodological frontier most relevant to a department on agentic decision systems. This frontier has advanced in four related directions. First, LLMs are increasingly used for measurement in text-rich social data, including annotation, coding, construct discovery, and downstream inference with imperfect surrogate labels (Ziems et al. 2024, Lee et al. 2024, Egami et al. 2023, Davidson & Karell 2025). Second, LLMs and agents are used as simulated respondents, consumers, economic agents, behaviorally similar chatbots, grounded individual simulations, and theory-grounded social agents (Argyle et al. 2023, Goli & Singh 2024, Filippas et al. 2023, Mei et al. 2024, Park et al. 2024, Manning & Horton 2026). Third, AI systems are being explored as collaborators in hypothesis generation, theory development, and automated social-science workflows (Grossmann et al. 2023, Bail 2024, Manning et al. 2024). Fourth, a growing methodological literature warns that these uses require validation against human behavior, domain evidence, and explicit causal designs, especially when simulated agents stand in for people, firms, or organizations (Gui & Toubia 2023, Abdurahman et al. 2025).

These developments reveal a substantial opening for decision science, precisely because the field's core problems differ from settings where agent-driven research has already found clearer routines. In many natural-science applications, agents search physical, chemical, biological, or engineering design spaces in which objectives, constraints, instruments, and validation pipelines are comparatively formalized. In computer science, fast progress is supported by benchmarks, code execution, synthetic environments, and rapid iteration around model capabilities. Decision science faces a less settled but especially important frontier: decisions are made

by human and artificial actors who adapt to one another, pursue plural and sometimes contested objectives, operate inside organizations and markets, and are governed by accountability regimes. This makes the opportunity larger, not smaller. Agents can help researchers explore decision architectures that are too high-dimensional for manual theorizing, while this department's framework can discipline that exploration through actors, flows, moderators, architectures, and field-to-model mapping.

For this reason, the department especially welcomes agent-driven research that uses agents to open new decision-science questions rather than only to automate familiar research tasks. Consider a researcher studying how content moderation architectures shape misinformation dynamics on a large platform. The relevant design space is too large to enumerate manually: escalation thresholds, reviewer capacity, appeal rights, confidence-score cutoffs, feedback timing, and moderator decision rules may interact in nonobvious ways. An agent-driven approach could deploy simulated user agents with heterogeneous susceptibility to misinformation and simulated moderation agents with different decision rules, allowing the researcher to identify which architectural features produce the most consequential outcomes before committing to costly field or lab interventions. In this mode, agents serve as a meta-architectural tool. They explore the design space systematically, while human researchers define the objective, interpret emergent regularities, and translate the patterns into targeted hypotheses about governance, human oversight, and platform-level outcomes.

The same features that make decision science a promising frontier also make agent-driven research difficult. First, researchers must specify objectives carefully: when agents explore on our behalf, should they optimize for prediction accuracy, robustness, novelty, interpretability, social welfare, scientific insight, or some combination of these goals? Second, agent-driven exploration requires alignment and governance. Human values, theoretical priors, and ethical constraints must be embedded in the discovery process, especially when simulated actors stand in for people, firms, or organizations affected by the resulting research. Third, generated insights require epistemic discipline. Simulated actors may encode invalid behavioral assumptions, open-ended search may produce plausible but spurious regularities, and findings may fail when identities, incentives, information, or governance structures change. Agent-driven research should therefore complement, not replace, traditional empirical work. We encourage multi-method designs in which agents search broadly and reveal candidate regularities, while field studies, lab experiments, simulations, and hybrid designs test whether those regularities survive measurement, institutional context, and causal scrutiny.

3.5 | Editorial preferences and fit criteria

This department welcomes research that advances understanding of agentic decision systems: environments in which

TABLE 1 Illustrative contribution contrasts for in-scope submissions.

Route	Meaningful contribution	Likely incremental contribution
Architectural	Identifies or tests a new decision architecture, or a theoretically important architectural parameter, such as routing rules, escalation thresholds, decision rights, feedback loops, or moderators, that changes how humans and agents generate decisions.	Moves a familiar architecture to a new domain without showing which actors, flows, moderators, or design parameters work differently.
Mechanistic	Identifies how agentic autonomy, learning, memory, tool use, or coordination changes reasoning, adaptation, conflict, trust, or outcome formation.	Imports a familiar mechanism from traditional decision science without showing why the mechanism operates differently when agents act, learn, coordinate, or delegate.
Governance	Shows how agentic systems change oversight, accountability, override rights, incentives, or risk thresholds as agents act, learn, coordinate, or delegate.	Applies familiar governance concepts to an agentic setting without identifying the new authority, risk, or accountability problem created by agentic behavior.
Methodological	Develops field-to-model mappings, experiments, simulations, agent-driven exploration, or validation strategies that make autonomous, adaptive, multi-actor processes observable and comparable.	Varies agentic features, such as autonomy, memory, or prompts, without improving identification, measurement, comparison, or validation.

autonomous AI agents and humans jointly generate, execute, and evaluate decisions over time. Central to our scope is the study of systems where agents exhibit autonomy, learning or adaptation, and coordination, and where decision outcomes emerge from interactions among multiple actors rather than from isolated human choice or static algorithmic output.

We *prioritize* work in which AI agents are integral decision participants and decisions emerge from structured interaction among agents and humans. Submissions may focus on (i) the autonomy, learning, and coordination of agents; (ii) the roles humans play in defining objectives, articulating values, interpreting context, adjudicating trade-offs, maintaining accountability, and guiding or constraining agentic processes; or (iii) the design of collaborative decision architectures that organize how humans and agents jointly produce outcomes. In all cases, the contribution should show how these elements meaningfully shape the decision environment, organizational processes, or system-level outcomes. Fitting the department's scope is a necessary but not sufficient condition for a strong contribution. Once a submission is within scope, we evaluate contribution by what it teaches about agentic decision systems, not merely by whether it applies an agentic system to a new domain. Table 1 translates the framework and methodological agenda developed above into illustrative contribution criteria. The table is not a checklist. Instead, it helps authors ask what becomes newly visible because the system is agentic: a decision architecture; a mechanism; a governance problem; or a method for studying autonomous, adaptive, multi-actor processes.

The key distinction is that domain novelty alone is not a scientific contribution. A study that deploys a familiar architecture in healthcare rather than finance and finds the same mechanism can provide useful replication, but a strong department contribution should explain what the replication teaches about portability, robustness, or boundary condi-

tions. A study that shows why the mechanism differs in healthcare, for example because the moderator regime is risk-triggered rather than efficiency-driven, contributes by identifying a boundary condition for generalization. Conversely, a study may contribute without a new domain if it reveals a previously unrecognized control signal, a feedback loop between human judgment and agent adaptation, or a meta-architectural design that makes dynamic agentic systems empirically tractable.

This department *does not* consider studies in which AI is treated primarily as a technical artifact or as a passive tool rather than as a meaningful decision participant. This includes the following: (1) Purely design-science or performance-driven work that focuses mainly on improving the technical capabilities, efficiency, or accuracy of agentic AI systems without engaging with how such systems change decision-making, organizational processes, or system-level outcomes. (2) Traditional, fixed, and fully predictable AI, including symbolic or rule-based systems whose behavior is entirely specified *ex ante*, or static analytical and optimization tools that do not learn, adapt, or interact over time.

We also *do not* consider studies that conceptualize human–AI interaction as isolated, one-shot, or purely dyadic exchanges, or that treat AI merely as an instrumental tool within such settings, where the AI neither exercises meaningful autonomy, coordinates with other agents, nor participates in a broader decision architecture. In these cases, the AI agents can typically be replaced by a fixed script, prompt, or heuristic, or a conventional decision aid, without altering the underlying decision logic, which limits the relevance of the contribution for understanding agentic or human–agent systems.

As a *practical heuristic*: Submissions are well aligned when AI agents are constitutive of the decision system, such that removing or substantially weakening them would

alter the system's architecture, dynamics, or outcomes. Submissions are misaligned when AI agents play a largely instrumental or technical role and could be replaced by a static rule, script, prompt, or conventional decision-support tool without changing the underlying decision logic.

4 | CONCLUDING REMARKS

The rise of agentic AI gives decision science a clear agenda: understand decision systems in which humans and autonomous agents jointly act, learn, and govern outcomes. This department aims to create a shared intellectual space where researchers can examine human–agent collaboration, rethink core assumptions, and develop a common conceptual language. We invite the community to pursue multi-method and multi-context research that explains when agentic systems improve decisions, when they create new risks, and how organizations can design them responsibly.

ORCID

Yingjie Zhang  <https://orcid.org/0000-0003-0745-2563>

REFERENCES

- Abdurahman, S., Salkhordeh Ziabari, A., Moore, A. K., Bartels, D.M., & Dehghani, M. (2025). A primer for evaluating large language models in social-science research. *Advances in Methods and Practices in Psychological Science*, 8(2).
- Adamopoulos, P., Ghose, A., & Tuzhilin, A. (2021). Heterogeneous demand effects of recommendation strategies in a mobile application: evidence from econometric models and machine-learning instruments. *MIS Quarterly*, 46 (1), 101–150.
- Adomavicius, G. & Yang, M. (2022). Integrating behavioral, economic, and technical insights to understand and address algorithmic bias: a human-centric perspective. *ACM Transactions on Management Information Systems (TMIS)*, 13(3), 1–27.
- Alavi, M. (2024). Research: how different fields are using GenAI to redefine roles. *Harvard Business Review*. <https://hbr.org/2024/03/research-how-different-fields-are-using-genai-to-redefine-roles>
- Argyle, L.P., Busby, E.C., Fulda, N., Gubler, J.R., Rytting, C. & Wingate, D. (2023). Out of one, many: using language models to simulate human samples. *Political Analysis*, 31(3), 337–351.
- Arora, S. & Doshi, P. (2021). A survey of inverse reinforcement learning: challenges, methods and progress. *Artificial Intelligence*, 297, 103500.
- Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, 29(3), 3–30.
- Bail, C. A. (2024). Can generative AI improve social science?. *Proceedings of the National Academy of Sciences*, 121(21), e2314021121.
- Baird, A. & Maruping, L. M. (2021). The next generation of research on is use: a theoretical framework of delegation to and from agentic artifacts. *MIS Quarterly*, 45(1), 315–341.
- Bayer, R.-C. & Renou, L. (2024). Interacting with man or machine: when do humans reason better? *Management Science*, 62(605).
- Bhargava, H. K., Brown, S., Ghose, A., Gupta, A., Leidner, D. & Wu, D. (2025). Exploring generative AI's impact on research: perspectives from senior scholars in management information systems. *ACM Transactions on Management Information Systems*, 16(2), 1–9.
- Bhatia, S., van Baal, S. T., Wang, F. & Walasek, L. (2025). Computational analysis of 100 k choice dilemmas: decision attributes, trade-off structures, and model-based prediction. *Proceedings of the National Academy of Sciences*, 122(17), e2406489122.
- Bi, X., Gupta, A. & Yang, M. (2024). Understanding partnership formation and repeated contributions in federated learning: An analytical investigation. *Management Science*, 70(8), 4974–4994.
- Brynjolfsson, E., Li, D. & Raymond, L. R. (2023). *Generative AI at work*. National Bureau of Economic Research.
- Brynjolfsson, E. & Mitchell, T. (2017). What can machine learning do? workforce implications. *Science*, 358(6370), 1530–1534.
- Brynjolfsson, E., Rock, D. & Syverson, C. (2021). The productivity j-curve: how intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics*, 13(1), 333–372.
- Cao, Z., Li, M. & Pavlou, P. A. (2024). AI in business research. *Decision Sciences*, 55(6), 518–532.
- Chen, Z. & Chan, J. (2024). Large language model in creative work: the role of collaboration modality and user expertise. *Management Science*, 70(12), 9101–9117.
- Choudhary, V., Marchetti, A., Shrestha, Y. R. & Puranam, P. (2025). Human-AI ensembles: when can they work?. *Journal of Management*, 51(2), 536–569.
- Davidson, T. & Karell, D. (2025). Integrating generative artificial intelligence into social science research: measurement, prompting, and simulation. *Sociological Methods & Research*, 54(3), 775–793.
- De Freitas, J., Nave, G. & Puntoni, S. (2025). Ideation with generative AI—in consumer research and beyond. *Journal of Consumer Research*, 52(1), 18–31.
- De Freitas, J., Oğuz-Uğuralp, Z., Uğuralp, A. K. & Puntoni, S. (2025). AI companions reduce loneliness. *Journal of Consumer Research*, 52(6), 1126–1148.
- 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), 307–337.
- Dietvorst, B. J., Simmons, J. P. & Massey, C. (2018). Overcoming algorithm aversion: people will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64(3), 1155–1170.
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350–383.
- Egami, N., Hinck, M., Stewart, B. M. & Wei, H. (2023). Using imperfect surrogates for downstream inference: design-based supervised learning for social science applications of large language models. *Advances in Neural Information Processing Systems*, 36, 68589–68601.
- Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 85–104.
- Feuerriegel, S., Shrestha, Y. R., von Krogh, G. & Zhang, C. (2022). Bringing artificial intelligence to business management. *Nature Machine Intelligence*, 4(7), 611–613.
- Filippas, A., Horton, J. J. & Manning, B. S. (2023). Large language models as simulated economic agents: what can we learn from homo silicus?. Working Paper 31 122, National Bureau of Economic Research.
- Fügener, A., Walzner, D. D. & Gupta, A. (2025). Roles of artificial intelligence in collaboration with humans: automation, augmentation, and the future of work. *Management Science*, 72(1), 538–557.
- Gao, S., Fang, A., Huang, Y., Giunchiglia, V., Noori, A., Schwarz, J. R., Ektefaie, Y., Kondic, J. & Zitnik, M. (2024). Empowering biomedical discovery with AI agents. *Cell*, 187(22), 6125–6151.
- Goldratt, E. M. (1990). *Theory of constraints*. North River Croton-on-Hudson.
- Goli, A. & Singh, A. (2024). Frontiers: Can large language models capture human preferences? *Marketing Science*, 43(4), 709–722.
- Gridach, M., Nanavati, J., Abidine, K. Z. E., Mendes, L. & Mack, C. (2025). Agentic AI for scientific discovery: a survey of progress, challenges, and future directions. *arXiv*. <https://doi.org/10.48550/arXiv.2503.08979>.
- Grossmann, I., Feinberg, M., Parker, D. C., Christakis, N. A., Tetlock, P. E. & Cunningham, W. A. (2023). AI and the transformation of social science research. *Science*, 380(6650), 1108–1109.
- Grote, G., Parker, S. K. & Crowston, K. (2026). Taming artificial intelligence: a theory of control-accountability alignment among AI developers and users. *Academy of Management Review*, 51(2), 278–299.

- Gui, G. & Toubia, O. (2023). The challenge of using llms to simulate human behavior: a causal inference perspective. *arXiv*. <https://doi.org/10.48550/arXiv.2312.15524>.
- Gupta, S., Rikhtehgar Berenji, H., Shukla, M. & Murthy, N. N. (2023). Opportunities in farming research from an operations management perspective. *Production and Operations Management*, 32(6), 1577–1596.
- Han, B. R., Chu, L. Y., Sun, T. & Wu, L. (2025). Commercializing the package flow: cross-sampling physical products through e-commerce warehouses. *Management Science*. 2025. <https://doi.org/10.1287/mnsc.2020.00902>
- Han, B. R., Sun, T., Chu, L. Y. & Wu, L. (2024). Connecting customers and merchants offline: experimental evidence from the commercialization of last-mile stations at alibaba. *MIS Quarterly*, 48(3), 1157–1192.
- Harrison, G. W. & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4), 1009–1055.
- IBM. (2025a). *AI agents in marketing*. Available at. <https://www.ibm.com/think/topics/ai-agents-in-marketing> [Accessed 30th December 2025].
- IBM. (2025b). *What is agentic AI*. Available at. <https://www.ibm.com/think/topics/agentic-ai> [Accessed 8th January, 2026].
- Ivanov, D., Tang, C. S., Dolgui, A., Battini, D. & Das, A. (2021). Researchers' perspectives on industry 4.0: multi-disciplinary analysis and opportunities for operations management. *International Journal of Production Research*, 59(7), 2055–2078.
- Jia, N., Luo, X., Fang, Z. & Liao, C. (2024). When and how artificial intelligence augments employee creativity. *Academy of Management Journal*, 67(1), 5–32.
- Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *The Annals of Mathematical Statistics*, 338–354.
- Köbis, N., Rahwan, Z., Rilla, R., Supriyatno, B. I., Bersch, C., Ajaj, T., Bonnefon, J.-F. & Rahwan, I. (2025). Delegation to artificial intelligence can increase dishonest behaviour. *Nature*, 646, 126–134.
- Koster, R., Balaguer, J., Tacchetti, A., Weinstein, A., Zhu, T., Hauser, O., et al., (2022). Human-centred mechanism design with democratic AI. *Nature Human Behaviour*, 6(10), 1398–1407.
- Lebovitz, S., Lifshitz-Assaf, H. & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organization Science*.
- Lee, D. D., Cheng, Z. Z., Mao, C. & Manzoor, E. (2025). Guided diverse concept miner (GDCM): Uncovering relevant constructs for managerial insights from text. *Information Systems Research*, 36(1), 370–393.
- Leng, Y. (2024). Can LLMS mimic human-like mental accounting and behavioral biases?. In: *Proceedings of the 25th ACM conference on economics and computation*, 8–11 July 2004 (pp. 581–581).
- Li, M. & Bitterly, T. B. (2024). How perceived lack of benevolence harmstrust of artificial intelligence management. *Journal of Applied Psychology*, 109(11), 1794–1816.
- Li, Y., Zhao, R., Li, X. & Choi, T.-M. (2024). Fine-tuning of artificial intelligence managers' logic in a supply chain with competing retailers. *Decision Sciences*, 55(6), 639–652.
- Lin, H., Czarnek, G., Lewis, B., White, J. P., Berinsky, A. J., Costello, T., Pennycook, G. & Rand, D. G. (2025). Persuading voters using human-artificial intelligence dialogues. *Nature*, 648, 394–401.
- Lou, B., Lu, T., Raghu, T. & Zhang, Y. (2025). Unraveling human-AI teaming: A review and outlook. *arXiv*. <https://doi.org/10.48550/arXiv.2504.05755>.
- Lu, T. & Zhang, Y. (2025). 1+ 1 > 2? information, humans, and machines. *Information Systems Research*, 36(1), 394–418.
- Ludwig, J. & Mullainathan, S. (2024). Machine learning as a tool for hypothesis generation. *The Quarterly Journal of Economics*, 139(2), 751–827.
- Luo, X., Tong, S., Fang, Z. & Qu, Z. (2019). Frontiers: machines vs. humans: the impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science*, 38(6), 937–947.
- Madras, D., Pitassi, T. & Zemel, R. (2018). Predict responsibly: improving fairness and accuracy by learning to defer. *Advances in Neural Information Processing Systems*, 31, 6150–6160.
- Manning, B. S. & Horton, J. J. (2026). General social agents. Working Paper 34937, National Bureau of Economic Research.
- Manning, B. S., Zhu, K. & Horton, J. J. (2024). Automated social science: Language models as scientist and subjects. Working Paper 32381, National Bureau of Economic Research.
- McKinsey. (2024). *Why AI agents are the next frontier of generative AI?*. Available at. <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-agents-are-the-next-frontier-of-generative-ai> [Accessed 30th December 2025].
- McKnight, D. H., Choudhury, V. & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: an integrative typology. *Information systems research*, 13(3), 334–359.
- Mei, Q., Xie, Y., Yuan, W., & Jackson M. O., (2024). A turing test of whether ai chatbots are behaviorally similar to humans. *Proceedings of the National Academy of Sciences*, 121(9), e2313925121.
- Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 17(4), 460–469.
- Özkan-Seely, G. F., Hall, D. C. & Hutchison-Krupat, J. (2024). Search for the best alternative: an experimental approach. *Decision Sciences*, 55(1), 57–67.
- Park, J. S., Zou, C. Q., Shaw, A., Hill, B. M., Cai, C. J., Morris, M. R., Willer, R., Liang, P. & Bernstein, M. S. (2024). Generative agent simulations of 1,000 people. *arXiv*. <https://doi.org/10.48550/arXiv.2411.10109>.
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., et al. (2019). Machine behaviour. *Nature*, 568(7753), 477–486.
- Revilla, E., Saenz, M. J., Seifter, M. & Ma, Y. (2023). Human-artificial intelligence collaboration in prediction: a field experiment in the retail industry. *Journal of Management Information Systems*, 40(4), 1071–1098.
- Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of personality*.
- Shan, G. & Qiu, L. (2025). Examining the impact of generative AI on users' voluntary knowledge contribution: Evidence from a natural experiment on stack overflow. *Information Systems Research*.
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99–118.
- Snyder, C., Keppler, S. & Leider, S. (2026). Algorithm reliance: Fast and slow. *Management Science* 72(1), 368–385.
- Srinivasan, R. & Sarial-Abi, G. (2021). When algorithms fail: consumers' responses to brand harm crises caused by algorithm errors. *Journal of Marketing*, 85(5), 74–91.
- Steyvers, M. & Kumar, A. (2024). Three challenges for AI-assisted decision-making. *Perspectives on Psychological Science*, 19(5), 722–734.
- Sun, C., Adamopoulos, P., Ghose, A. & Luo, X. (2021). Predicting stages in the consumer path-purchase journey: an omnichannel deep-learning model. *SSRN*. <https://doi.org/10.2139/ssrn.3814630>.
- Sun, T. & Taylor, S. J. (2020). Displaying things in common to encourage friendship formation: a large randomized field experiment. *Quantitative Marketing and Economics*, 18(3), 237–271.
- Sun, T., Yuan, Z., Li, C., Zhang, K. & Xu, J. (2024). The value of personal data in internet commerce: a high-stakes field experiment on data regulation policy. *Management Science*, 70(4), 2645–2660.
- Susarla, A., Gopal, R., Thatcher, J. B. & Sarker, S. (2023). The janus effect of generative AI: charting the path for responsible conduct of scholarly activities in information systems. *Information Systems Research*, 34(2), 399–408.
- Swanson, K., Wu, W., Bulaong, N. L., Pak, J. E. & Zou, J. (2025). The virtual lab of AI agents designs new sars-cov-2 nanobodies. *Nature*, 646(8085), 716–723.
- Tambe, P. B. (2025). Reskilling the workforce for AI: Domain expertise and algorithmic literacy. *Management Scienc*, 72(1), 515–537.
- Tang, C. S. (2025). Managing supply chain risks in the AI era. In: *AI in Supply Chains: Perspectives from Global Thought Leaders*, Yale University Press, New Haven, CTpp. 13–20.
- Thaler, R. H. (1988). Anomalies: the ultimatum game. *Journal of Economic Perspectives*, 2(4), 195–206.

- Thaler, R. H. & Sunstein, C. R. (2021). *Nudge: The final edition*. London: Penguin.
- Tuckman, B. W. (1965). Developmental sequence in small groups. *Psychological Bulletin*, 63(6), 384.
- Wang, H., Zhang, Y. & Lu, T. (2026). The power of disagreement: A field experiment to investigate human–algorithm collaboration in loan evaluations. *Management Science*, 72(1), 96–118.
- Wang, L., Huang, N., He, Y., Liu, D., Guo, X., Sun, Y. & Chen, G. (2025). Artificial intelligence (AI) assistant in online shopping: a randomized field experiment on a livestream selling platform. *Information Systems Research*.
- Wang, Q., Wang, T., Tang, Z., Li, Q., Chen, N., Liang, J. & He, B. (2025). Megaagent: A large-scale autonomous LLM-based multi-agent system without predefined sops. In *Findings of the 2025 Conference of the Association for Computational Linguistics*. Association for Computational Linguistics, Vienna, Austria, 4998–5036. <https://doi.org/10.18653/v1/2025.findings-acl.259>
- Wang, W., Pei, S. & Sun, T. (2023). Unraveling generative AI from a human intelligence perspective: A battery of experiments. *SSRN*. <https://ssrn.com/abstract=4543351>.
- Wang, W., Yang, M. & Sun, T. (2023). Human-AI co-creation in product ideation: the dual view of quality and diversity. *Available at SSRN 4668241*.
- Wang, X., Feng, C. & Sun, T. (2025). AI spillover is different: flat and lean firms as engines of AI diffusion and productivity gain. *arXiv preprint arXiv:2511.02099*.
- Weick, K. E. & Weick, K. E. (1995). *Sensemaking in organizations*, vol. 3. California: Sage publications.
- Womack, J. P., Jones, D. T. & Roos, D. (2007). *The machine that changed the world: the story of lean production—Toyota's secret weapon in the global car wars that is now revolutionizing world industry*. New York: Simon and Schuster.
- Wu, Z., Yu, C., Chen, C., Hao, J. & Zhuo, H. H. (2023). Models as agents: Optimizing multi-step predictions of interactive local models in model-based multi-agent reinforcement learning. In: *Proceedings of the AAAI conference on artificial intelligence*, Washington, DC, USA: AAAI Press, volume 37, 7–14 Feb 2023, Washington, D.C., USA. pp. 10435–10443.
- Xu, X., Wang, Q., Liu, Y. & Qiu, L. (2025). Social audience size as a reference point: evidence from a field experiment. *Management Science*.
- Yu, Y., Xue, W., Jia, L. & Tan, Y. (2025). When emotion AI meets strategic users. *Management Science*.
- Zhang, S., Yu, J., Xu, X., Yin, C., Lu, Y., Yao, B., et al. (2020). Rethinking human-AI collaboration in complex medical decision making: A case study in sepsis diagnosis. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, (pp. 1–18).
- Zhang, W. & Ram, S. (2025). A comprehensive analysis of triggers and risk factors for asthma based on machine learning and large heterogeneous data sources. *MIS Quarterly*, 44(1).
- Zhang, Y., Zheng, Z. & Gu, B. (2024). Omnificence or differentiation? an empirical study of knowledge structure and career development of it workers. *Information Systems Research*, 36(2), 1129–1146.
- Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z. & Yang, D. (2024). Can large language models transform computational social science?. *Computational Linguistics*, 50(1), 237–291.
- Zöller, N., Berger, J., Lin, I., Fu, N., Komarneni, J., Barabucci, G., et al., (2025). Human-ai collectives most accurately diagnose clinical vignettes. *Proceedings of the National Academy of Sciences*, 122(24), e2426153122.

How to cite this article: Zhang, Y., & Sun, T. (2026) Decision science in the agentic era: Research frontiers on agentic AI and human-agent collaboration. *Decision Sciences*, 1–15. <https://doi.org/10.1111/dec.70037>

AUTHOR BIOGRAPHIES

Yingjie Zhang Guanghua School of Management, Peking University Beijing, 10087 China.

Tianshu Sun Cheung Kong Graduate School of Business Oriental Plaza, Tower E3, 3F One East Chang An Avenue Beijing 100738, China.