Orchestration and Agency in Multi-Agent Systems: Preserving Human Capability in Agentic Automation

ALBERTO VARONE and FRANCESCO BOLICI, University of Cassino and Southern Lazio, Italy
KEVIN CROWSTON, Syracuse University, USA

Multi-agent AI systems promise autonomous problem-solving through coordinated specialized agents. We argue these systems amplify the orchestration requirements and verification challenges inherent to AI use, creating acute risks for human agency and skill development. Synthesizing Information Processing Theory and sensemaking perspectives, we identify three critical challenges: attribution and transparency for verification, the orchestration manager paradox and compound deskilling risks. We describe three divergent skill trajectories—novice compression, intermediate drift, expert expansion—and conclude with design recommendations and research questions.

CCS Concepts: • **Computing methodologies** → **Artificial intelligence**; **Multi-agent systems**; • **Human-centered computing** → *Collaborative and social computing theory, concepts and paradigms.*

Additional Key Words and Phrases: Organizational information processing, multi-agent systems, orchestration, generative artificial intelligence

ACM Reference Format:
Alberto Varone, Francesco Bolici, and Kevin Crowston. 2026. Orchestration and Agency in Multi-Agent Systems: Preserving Human Capability in Agentic Automation. In *Proceedings of AutomationXP26 Workshop of the 2026 CHI Conference on Human Factors in Computing Systems, April 14, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 7 pages.

1 Introduction: The Challenge of Orchestration

Despite significant investments in artificial intelligence (AI) systems, many organizations have yet to achieve meaningful returns or full integration into workflows [15]. The lack of progress stems in part from a prevalent techno-centric view of AI implementation that treats it primarily as a technological substitution for how tasks are performed rather than an opportunity for deeper organizational transformation [3, 25]. The techno-centric perspective neglects key organizational dimensions, including task interdependencies and associated coordination needs [1, 22], leaving these factors to human oversight. Consequently, despite organizations investing significant resources in AI implementation, they face significant challenges in achieving effective impact beyond isolated tasks.

The challenge of integrating AI into long-standing work practices becomes more complex as we move from single-agent systems to agentic AI architectures in which multiple specialized agents form a network and coordinate to solve complex problems with minimal human involvement. For instance, software development may involve different agents addressing planning, code development and testing; writing, agents to handle research, outlining, drafting and revision. We argue that multi-agent systems amplify both the orchestration requirements and the verification challenges that

Authors' Contact Information: Alberto Varone, alberto.varone@unicas.it; Francesco Bolici, f.bolici@unicas.it, University of Cassino and Southern Lazio, Cassino, Italy; Kevin Crowston, Syracuse University, Syracuse, NY, USA, crowston@syr.edu.

Manuscript submitted to ACM

characterize work with AI, creating new and more acute risks for human agency and skill development. While agentic AI promises to reduce human intervention, this apparent autonomy obscures critical questions: Who orchestrates the orchestrators? How do humans maintain meaningful agency when coordination happens between machines? What happens to skill development when learning opportunities are distributed across an opaque agent network?

To address these questions, we synthesize insights from Information Processing Theory [11, 21] applied to GenAI orchestration and sensemaking theory [14, 23] applied to AI-mediated skill development. Our central claim is that multi-agent systems do not eliminate coordination challenges but rather redistribute them in ways that make human orchestration more critical yet simultaneously harder to execute effectively.

2 Multi-Agent Orchestration: Layered Complexity

Information Processing Theory points out that organizations face dual information-processing requirements when accomplishing organizational tasks: the direct management of information needed for task execution and the coordination of information flows among interdependent tasks [21]. AI systems can contribute both Processing Capacity (PC), i.e., their ability to perform tasks, and Orchestration Capability (OC), i.e., their ability to arrange information flows that manage task interdependencies [4]. OC encompasses inbound orchestration (identifying, accessing and retrieving necessary information) and outbound orchestration (capturing, formatting and disseminating outputs to dependent downstream tasks). We suggest that a system may posses ample PC, meaning it can successfully perform tasks, but orchestration is often left to the human user who has to martial appropriate inputs to the system and verify outputs and incorporate them into the final work product. We therefore view human-AI collaboration as an orchestration cycle with four phases through which the human manages the AI: framing the problem, delegating through prompting, interpreting and verifying system outputs and integrating them into a final work product. For instance, a customer service representative using an AI chatbot assistant must frame customer issues appropriately, prompt for responses, interpret the AI's suggestions in context of company policies and integrate approved responses into their conversation. This cycle is manageable because there is one system to verify and one set of outputs to check.

In multi-agent systems, however, this cycle becomes distributed and recursive. Framing occurs at the network and individual agent levels. Delegation involves configuring agents and agent interactions. Interpretation requires understanding agent outputs and how they interact. Most critically, verification must operate at multiple levels: verifying individual agents, inter-agent flows and overall solution coherence. The question shifts from "Is this output correct?" to "Is this network of coordinated outputs correct? And if not, can I trace errors to their source?" As a result, multi-agent systems create layered orchestration needs beyond single-agent deployments: intra-agent orchestration (each agent managing its own inputs/outputs), inter-agent orchestration (agents coordinating with one another) and human-agent network orchestration (humans overseeing coordination patterns across the entire agent network). If multiple AI agents possess only PC without corresponding OC, significant process-level requirements will remain unaddressed, creating bottlenecks and failures that may be hidden within inter-agent handoffs [4]. We note that research on AI's potential role in coordination remains less explored compared to its role in communication or collaboration at the task level.

3 Three Critical Challenges

Layered orchestration demands create challenges for human-agent collaboration. While multi-agent systems promise reduced human intervention, they simultaneously raise fundamental questions about human agency, verification capabilities and skill development. We identify three critical challenges that must be addressed to preserve meaningful human involvement.

3.1 Attribution and Transparency for Multi-Agent Verification

First, drawing on sensemaking perspectives, we identify verification as the most consequential phase in the orchestration cycle: the point at which users pause to interrogate AI-generated material, compare it with their understanding and decide whether to accept, revise or reject it. When verification weakens, the cycle collapses into passive use; when it strengthens, interaction becomes genuine sensemaking [14, 23]. Individuals who interrogate and reinterpret system outputs develop richer mental models and retain greater agency, whereas those who rely uncritically on AI tend to plateau [13, 20].

However, verification depends fundamentally on attribution: users must know what was generated, by which system and based on which inputs. These questions are harder to answer in a multi-agent network. Consider a scenario where Agent A gathers data, Agent B analyzes it, Agent C generates recommendations and Agent D formats output. If the final deliverable contains an error, which agent is responsible? Without clear attribution mechanisms, users cannot effectively verify outputs because they cannot trace information flows backward. This attribution problem compounds the residual equivocality characterizing information exchange and coordination. In multi-agent systems, equivocality can accumulate across agent handoffs: each interface introduces potential for misinterpretation and ambiguities propagate through the network. Multi-agent opacity compounds verification difficulty, increasing the risk of uncritical reliance.

3.2 Human Agency: The Orchestration Manager Paradox

Second, prevailing discourse equates AI autonomy with its capacity to perform a specified task without direct human intervention. However, process-contextualized autonomy demands that GenAI systems possess not only high PC but also robust OC [4]. An AI might demonstrate impressive autonomy in executing its core function but introduce significant friction if it cannot also autonomously manage its information dependencies. This realization directly challenges the plug-and-play assumption that AI can seamlessly take on delegated tasks without disruption to the overall work system. Successful AI implementation instead necessitates diligent assessment and deliberate design of the AI's orchestration profile to address coordination requirements. In multi-agent contexts, this challenge intensifies: each agent may appear capable in isolation, but network-level autonomy requires agents to coordinate among themselves with minimal human intervention.

To manage these concerns, new human roles are emerging, e.g., Orchestration Managers or Human-AI System Integrators [4]. In multi-agent scenarios, humans not only oversee individual agents but actively manage information pipelines across agents, troubleshoot orchestration breakdowns and ensure seamless handoffs. This form of deployment creates a challenge: agentic systems supposedly reduce human involvement, yet effective deployment requires humans to manage agent coordination at higher abstraction levels. The question becomes not whether humans remain involved, but at which level they are able to exercise agency, verifying individual outputs (micro-level) or coordinating the agent network (macro-level). When agent-to-agent coordination is opaque, humans may be unable to exercise either form of agency effectively.

3.3 Skill Development: The Compound Deskilling Risk

Finally, AI reduces the cognitive effort required to produce acceptable outputs, particularly for less experienced users. This support improves performance in the short run but may weaken mechanisms through which deeper understanding develops [19]. Traditional skill development often relies on "learning by doing": cycles where workers confront ambiguity,

interpret feedback and internalize routines [2, 8, 9]. With AI, these cycles increasingly become "doing without learning" as AI handles the intermediate reasoning steps that functioned as engines of cognitive growth. Tasks are completed successfully, yet interpretive and diagnostic routines underpinning skill development are bypassed [7, 16, 17]. Multi-agent systems compound this risk by distributing learning opportunities across the network, each agent potentially removing cognitive work that historically supported skill development.

Three skill trajectories emerge: novice compression, intermediate drift and expert expansion:

- Novice compression: Novices lack the well-developed interpretive schemas needed to scrutinize AI outputs. Even when intending to verify, they may not recognize inconsistencies [20]. In multi-agent systems, novices must understand not just one agent's output but how multiple contributions fit together. For instance, a junior analyst may succeed in getting results from a pipeline in which data flows from extraction to pre-processing to analysis to visualization to presentation, but without developing a complete understanding, e.g., intuitions about how the pre-processing affects the results. Their use of AI can rapidly improve surface performance without deepening underlying competence, shrinking the range of tasks where they actively problem-solve rather than defer to AI.
- Intermediate drift: Users at intermediate skill levels possess emerging schemas allowing them to sense when output seems off, but their ability to diagnosis remains fragile. For them, AI can powerfully amplify learning if verification remains active: discrepancies become resources for interpretation and iterative sensemaking refines developing mental models [24]. However, intermediates may oscillate between accelerated learning and stagnation depending on how they interrogate outputs. For instance, reading agent-generated code may yield new insights about techniques or libraries, but in a time crunch, entire agent-generated modules might be accepted with minimal or no review. In multi-agent contexts, the complexity of verifying networked agents may exceed intermediate capacity, causing plateaus precisely when progression is needed.
- Expert expansion: Experts use AI to expand available cues. Strong mental models allow efficient verification and integration without losing interpretive control. For instance, a consultant could develop and compare multiple solution options, deepening their understanding of the situation. In multi-agent systems, experts can leverage network-level reasoning as catalyst for deeper representational refinement, treating agent coordination patterns as analysis objects rather than black boxes.

The verification bottleneck links cognitive, social and organizational dynamics, determining whether AI substitutes for thinking or catalyzes practice-based learning [6]. In multi-agent systems, this bottleneck intensifies: more outputs to verify, less understanding of agent interactions and reduced ability to trace errors. The key differentiator among users is not AI proficiency but "verification intent": the motivation and ability to continuously interrogate outputs rather than accepting them at face value [4]. This intent shapes whether individuals retain inquiry and feedback cycles enabling learning by doing or drift into patterns where performance improves but understanding stagnates.

4 Implications and Research Directions

Successful AI integration requires redesigning entire workflows to balance execution and orchestration. For multi-agent systems, this redesign must operate at multiple levels simultaneously.

Design principles: Multi-agent systems must surface the coordination patterns underlying the task execution, not just final outputs. Systems exposing reasoning steps, foregrounding exceptions and inviting intervention facilitate verification and preserve learning opportunities [12]. In multi-agent contexts, visibility means showing not just what each agent did but how agents coordinated. Designer can include explicit verification checkpoints at inter-agent handoffs

and develop the technical infrastructure needed to make agent contributions traceable, e.g., through version control, logging or explicit contribution tagging.

Organizational design: Organizations must recognize and support orchestration manager roles. These positions will require skills extending beyond checking factual accuracy to evaluating information processing efficacy: how well did agents consume inputs and prepare outputs for downstream use [4]? Furthermore, verification is shaped by the social and organizational environment [23]. Cultures that emphasize learning, critique and psychological safety encourage deeper engagement with AI outputs. Conversely, environments prioritizing speed and efficiency discourage verification, especially for novices fearing signals of incompetence. Team norms around documentation, peer review and reflective practice can also amplify or suppress verification behavior [14]. Universities face similar challenges as students circumvent the interpretive work curricula are designed to develop. Embedding explicit instruction in verification strategy is now essential. Finally, HR policies should preserve human worker access to the tasks that are crucial for continued learning through rotational assignments, manual-first exercises or workflows surfacing intermediate reasoning [10, 18], resisting the temptation to deploy learning-destroying automation for short-term gain.

Research questions: We end with several questions to guide future research. How can we measure OC at individual agent and network levels? For instance, what behavioral signatures indicate high vs. low inter-agent orchestration capability? Should we track handoff success rates, error propagation patterns or the extent of human intervention required? What attribution mechanisms support verification without overwhelming users? Should systems log every agent contribution? Provide "audit trails" on demand? Use visualization to show information flows? How do we balance transparency with cognitive load? When do multi-agent systems genuinely achieve process autonomy versus obscuring coordination challenges? What competencies distinguish effective orchestration managers from passive users?How can we empirically track skill trajectories in agentic environments?

5 Conclusion

Agentic AI represents a qualitative shift in the needed coordination structure, not merely more autonomous AI. The movement to multi-agent systems redistributes orchestration challenges in ways that make human orchestration more critical yet simultaneously harder to execute effectively. Success requires preserving human agency at critical sensemaking junctions [24]. Effective AI integration is about alignment: whether the work is driven by humans, machines or their interaction, orchestration capability is critical for achieving organizational fit [4]. Further, skill development requires both exposure to meaningful cues and opportunities to interpret and verify them [14, 23]. AI systems modify both elements, redistributing interpretive work between humans and machines. Whether this redistribution leads to accelerated learning or shallow proficiency depends on users' verification intent and the sociotechnical contexts enabling or suppressing interpretive engagement [5].

Returning to our opening question—who orchestrates the orchestrators?—the answer is not that humans can or should be removed from orchestration. Rather, as systems become more agentic, human orchestration must operate at higher levels of abstraction while remaining grounded in verifiable, traceable agent behaviors. The challenge is designing systems and organizations that preserve human agency where it matters most: at the sensemaking junctions where coordination patterns are established, verified and refined.

Generative AI Tool Use Disclosure

An LLM was used to prepare an initial draft of this abstract by condensing a longer unpublished paper written by the authors. The draft was then checked and extensively revised. The revision process included using an LLM to review the paper to identify weaknesses in the presentation. The authors take full responsibility for the paper's contents.

Acknowledgments

This work was partly funded by a grant from the Alfred P. Sloan Foundation, G-2025-79202.

References

[1] Ajay Agrawal, Joshua S. Gans, and Avi Goldfarb. 2024. Artificial intelligence adoption and system-wide change. *Economics & Management Strategy* 33 (2024), 327–337. doi:10.1111/jems.12521

[2] John R. Anderson. 1982. Acquisition of cognitive skill. *Psychological Review* 89, 4 (1982), 369. doi:10.1037/0033-295X.89.4.369

[3] Hind Benbya, Thomas H. Davenport, and Stella Pachidi. 2020. Special issue editorial: Artificial intelligence in organizations: Current state and future opportunities. *MIS Quarterly Executive* 19 (2020), article 4. https://aisel-aisnet-org.libezproxy2.syr.edu/misqe/vol19/iss4/4

[4] Franceso Bolici, Alberto Varone, and Gabriele Diana. 2025. Beyond AI automation: How replace, reinforce, reveal, and orchestrate modes align task interdependence and organizational design for effective AI implementation. In *Theorizing Data & AI Workshop*. Amsterdam, Netherlands.

[5] Marie-Claude Boudreau and Daniel Robey. 2005. Enacting integrated information technology: A human agency perspective. *Organization Science* 16, 1 (2005), 3–18. doi:10.1287/orsc.1040.0103

[6] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z Gajos. 2021. To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-computer Interaction* 5, CSCW1 (2021), 1–21. doi:10.1145/3449287

[7] Fabrizio Dell'Acqua, Edward McFowland, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani. 2023. Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. doi:10.2139/ssrn.4573321

[8] Stuart E. Dreyfus. 2004. The five-stage model of adult skill acquisition. *Bulletin of Science, Technology & Society* 24, 3 (2004), 177–181. doi:10.1177/0270467604264992

[9] K. Anders Ericsson, Ralf T. Krampe, and Clemens Tesch-Römer. 1993. The role of deliberate practice in the acquisition of expert performance. *Psychological Review* 100, 3 (1993), 363. doi:10.1037/0033-295X.100.3.363

[10] Samer Faraj, Stella Pachidi, and Karla Sayegh. 2018. Working and organizing in the age of the learning algorithm. *Information and Organization* 28, 1 (2018), 62–70. doi:10.1016/j.infoandorg.2018.02.005

[11] Jay R. Galbraith. 1974. Organization design: An information processing view. *Interfaces* 4 (1974), 28–36. doi:10.1287/inte.4.3.28

[12] Clemens Haußmann, Yogesh K. Dwivedi, Krishna Venkitachalam, and Michael D. Williams. 2012. A summary and review of Galbraith's organizational information processing theory. In *Information Systems Theory*, Yogesh K. Dwivedi, Michael Wade, and Scott L. Schneberger (Eds.). Springer, New York, NY, 71–93. doi:10.1007/978-1-4419-9707-4_5

[13] Maurice Jakesch, Jeffrey T. Hancock, and Mor Naaman. 2023. Human heuristics for AI-generated language are flawed. *Proceedings of the National Academy of Sciences* 120, 11 (2023), e2208839120. doi:10.1073/pnas.2208839120

[14] Sally Maitlis and Marlys Christianson. 2014. Sensemaking in organizations: Taking stock and moving forward. *Academy of Management Annals* 8, 1 (2014), 57–125. doi:10.5465/19416520.2014.873177

[15] Hannah Mayer, Lareina Yee, Michael Chui, and Roger Roberts. 2025. *Superagency in the workplace: Empowering people to unlock AI's full potential*. Technical Report. McKinsey & Company. https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/superagency-in-the-workplace-empowering-people-to-unlock-ais-full-potential-at-work

[16] Shakked Noy and Whitney Zhang. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* 381, 6654 (2023), 187–192. doi:10.1126/science.adh2586

[17] Sida Peng, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. 2023. The impact of AI on developer productivity: Evidence from GitHub Copilot. arXiv:arXiv:2302.06590

[18] Sebastian Raisch and Sebastian Krakowski. 2021. Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review* 46, 1 (2021), 192–210. doi:10.5465/amr.2018.0072

[19] Tapani Rinta-Kahila, Esko Penttinen, Antti Salovaara, Wael Soliman, and Joona Ruissalo. 2023. The vicious circles of skill erosion: A case study of cognitive automation. *Journal of the Association for Information Systems* 24, 5 (2023), 1378–1412. doi:10.17705/1jais.00829

[20] Sara Salimzadeh, Gaole He, and Ujwal Gadiraju. 2024. Dealing with uncertainty: Understanding the impact of prognostic versus diagnostic tasks on trust and reliance in human-AI decision making. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17. doi:10.1145/3613904.3641905

[21] Michael L. Tushman and David A. Nadler. 1978. Information processing as an integrating concept in organizational design. *Academy of Management Review* 3 (1978), 613–624. doi:10.5465/amr.1978.4305791

[22] Melissa A. Valentine, Amanda L. Pratt, Rebecca Hinds, and Michael S. Bernstein. 2024. The algorithm and the org chart: How algorithms can conflict with organizational structures. *Proceedings of the ACM on Human-Computer Interaction* 8 (2024), 364:1–364:31. Issue CSCW. doi:10.1145/3686903

[23] Karl E. Weick. 1995. *Sensemaking in Organizations.* Sage, Thousand Oaks, CA.

[24] Karl E. Weick, Kathleen M. Sutcliffe, and David Obstfeld. 2005. Organizing and the process of sensemaking. *Organization Science* 16, 4 (2005), 409–421. doi:10.1287/orsc.1050.0133

[25] Xinying Yu, Shi Xu, and Mark Ashton. 2022. Antecedents and outcomes of artificial intelligence adoption and application in the workplace: The socio-technical system theory perspective. *Information Technology & People* 36 (2022), 454–474. doi:10.1108/ITP-04-2021-0254