

Artificial Intelligence as an Epistemic Co-Agent: Redefining the Scientific Research Process and Organizational Practices

Aya Daruich

PhD Candidate, LEG Laboratory, Faculty of Legal, Economic and Social Sciences,

Sultan Moulay Slimane University, Morocco

ayadaruich09@gmail.com

Abstract—The increasing integration of artificial intelligence (AI) into scientific research is profoundly transforming the processes through which knowledge is produced and validated. While prior studies predominantly conceptualize AI as an instrumental tool, limited attention has been paid to its role as an epistemic actor in knowledge construction. Addressing this gap, this paper conceptualizes AI as an epistemic co-agent that actively participates in the scientific research process. Adopting a conceptual research design, the study develops an analytical framework characterizing the various forms of human-AI interaction and the emergence of hybrid intelligence configurations in scientific production, organized around an original three-configuration typology: the researcher as producer, collaborator, or supervisor. The paper further examines the organizational and managerial implications of AI integration, including transformations in research practices, evolving decision-making processes, and governance challenges. This study makes three primary contributions: (1) it introduces the concept of AI as an epistemic co-agent; (2) it develops a conceptual framework of human-AI epistemic collaboration; and (3) it outlines governance mechanisms for AI-augmented research environments.

Keywords—Artificial Intelligence, Epistemic Co-Agent, Epistemic Agency, Hybrid Intelligence, Knowledge Production, Research Governance, Human-AI Interaction, Organizational Practices, Sociomateriality.

I. INTRODUCTION

In February 2026, Google DeepMind's Aletheia agent produced an autonomous mathematical research paper—without any human intervention—calculating structure constants in arithmetic geometry [6]. Days later, the same team's AI Co-scientist independently proposed drug repurposing hypotheses that matched findings researchers had spent years developing [10]. These events signal a fundamental reconfiguration of the scientific research process.

Artificial intelligence has evolved from a computational tool to an active participant in knowledge production. Yet existing theoretical frameworks have struggled to keep pace. Most organizational and management studies still conceptualize AI as an instrument—sophisticated, but ultimately subordinate to human agency. This framing is increasingly inadequate.

This paper addresses that gap by conceptualizing AI as an *epistemic co-agent*: an entity that actively participates in knowledge construction alongside the researcher. Our central question is: *How does the integration of artificial intelligence as an epistemic co-agent reshape scientific research processes and organizational practices, and what are its implications for knowledge production and governance?*

Our contributions are threefold. First, we introduce and define the concept of epistemic co-agency as applied to AI. Second, we propose an original typology of human-AI epistemic configurations. Third, we derive organizational and governance implications. This paper adopts a theory-building conceptual approach [13], drawing on an integrative review across epistemology, information systems, organizational theory, and AI science. Following Cornelissen's (2017) guidance on conceptual theorizing in management research, we proceed by identifying a gap in existing conceptualizations of AI in science, introducing a novel construct—epistemic co-agency—and deriving a typology that structures future empirical inquiry through testable propositions.

II. LITERATURE REVIEW

A. AI in Scientific Research

The role of AI in scientific research has evolved considerably. Initially deployed for computational tasks—data processing, pattern recognition, statistical analysis—AI systems progressively moved into more cognitively demanding functions. Xu et al. [20] characterize this trajectory as AI becoming a powerful paradigm for

scientific research, capable of hypothesis generation, simulation, and cross-disciplinary synthesis. Google DeepMind's AI Co-scientist, built on Gemini 2.0, autonomously generates novel research hypotheses across biomedical domains [10], while the Genesis Mission aims to deploy similar systems across U.S. National Laboratories to accelerate scientific productivity [9].

B. Epistemic Agency and Knowledge Production

Epistemic agency concerns the capacity of an entity to actively participate in the production, validation, and revision of knowledge [3]. Alvarado [1] argues that AI systems qualify as epistemic technologies—uniquely designed to operate in inquiry contexts and manipulate epistemic content. Hauswald [11] further assesses conditions under which AI systems may be recognized as artificial epistemic authorities whose outputs carry legitimate knowledge claims. As Coeckelbergh [3] notes, AI does not merely provide more information; it reshapes the very mechanisms through which beliefs are formed and revised.

C. Hybrid Intelligence and Human-AI Configurations

The concept of hybrid intelligence, introduced by Dellermann et al. [4], defines a sociotechnical configuration where human and artificial intelligence combine to achieve outcomes superior to either alone. Hemmer et al. [12] demonstrate that human-AI collaboration in decision-making produces superior outcomes when capability asymmetries are leveraged. The key insight is that the configuration of collaboration—who leads, who validates, who decides—determines the quality of the epistemic outcome.

Recent work extends this perspective in two directions. First, Rughiniş et al. [19] conceptualize AI "at the knowledge gates" as a gatekeeping force that filters which claims achieve epistemic legitimacy. Second, Levy et al. [14] argue that human-AI collaborations require community-level standards of epistemic responsibility—what they term "accountability without authorship"—anticipating institutional arrangements in which researchers remain answerable for outputs they did not solely produce. Together, these contributions suggest that human-AI configurations cannot be characterized by task allocation alone; they require explicit theorization of epistemic roles and responsibilities, which this paper develops.

D. Organizational and Governance Implications

Frontiers in Research Metrics [7] documents a growing reproducibility crisis in AI-augmented science, driven by unstable models and opaque training procedures. Chen et al. [2] highlight how AI facilitates data fabrication threatening research integrity. NIST [16] identifies algorithmic bias as a systemic risk. GESDA's 2026 Science Breakthrough Radar identifies AI governance as the central challenge of the decade: scientific capability is accelerating faster than institutional readiness.

III. CONCEPTUALIZING AI AS AN EPISTEMIC CO-AGENT

A. From Tool to Co-Agent: A Conceptual Progression

We propose distinguishing three positions AI can occupy. As a **tool**, AI executes predefined instructions without contributing cognitively. As an **assistant**, AI accelerates human work but the epistemic initiative remains entirely human. As an **epistemic co-agent**, AI actively participates in the construction of knowledge: generating hypotheses, identifying non-obvious patterns, proposing research directions the human researcher had not envisioned. This distinction is not merely technical but epistemological: when AI generates a hypothesis that a researcher validates and builds upon, the knowledge produced is genuinely co-constructed.

This conceptualization resonates with sociomateriality theory [18], which holds that social and material elements are inseparably entangled in the production of organizational outcomes. Unlike related notions such as decision-support systems or algorithmic agency, the epistemic co-agent specifically emphasizes the active participation of AI in the generation, validation, and legitimation of scientific knowledge.

Three neighboring concepts must be distinguished to clarify the contribution. Hybrid intelligence [4] addresses the functional complementarity of human and artificial capabilities, focusing on task performance rather than on the epistemic status of outputs. Epistemic technologies [1] characterize AI as an instrument for inquiry but stop short of granting it participatory agency in knowledge claims. Sociomateriality [18] foregrounds the entanglement of social and material elements yet remains largely silent on how scientific validity is produced and attributed. Epistemic co-agency is ontologically distinct: it specifically designates AI's participation in the generation, validation, and legitimation of knowledge as accepted truth-claims within scientific communities.

B. Dimensions of Epistemic Co-Agency

We identify four core dimensions: (1) **Cognitive contribution**—AI's capacity to generate novel epistemic content; (2) **Functional autonomy**—the degree to which AI operates without constant human instruction; (3) **Human-AI interaction quality**—the nature of the iterative dialogue between researcher and system; (4) **Influence on knowledge validation**—AI's role in shaping what counts as credible scientific output.

C. A Typology of Human-AI Epistemic Configurations

Drawing on Dellermann et al.'s (2019) hybrid intelligence framework, we propose a three-configuration typology grounded in two axes: the degree of AI functional autonomy, and the epistemic role assumed by the human researcher. As AI autonomy increases—from narrow AI to frontier generative systems to fully agentic AI—the researcher's role necessarily evolves from producer to collaborator to supervisor.

Configuration 1—Human as Producer: The researcher retains full epistemic initiative. AI is a sophisticated tool or assistant. Knowledge production is human-led; AI contributes execution, not ideation.

Configuration 2—Human as Collaborator: Researcher and AI engage in genuine epistemic dialogue. Knowledge is co-constructed through iterative exchange, as in systems like the AI Co-scientist where researchers direct inquiry while AI generates and evaluates hypotheses.

Configuration 3—Human as Supervisor: AI assumes the primary epistemic role—generating, verifying, and formalizing knowledge outputs. The researcher evaluates and validates. This configuration is exemplified by Aletheia [6] and is the most epistemically risky: circular validation increases when AI both produces and verifies knowledge; responsibility becomes diffuse; and fully autonomous AI is structurally incapable of the paradigmatic rupture that drives scientific progress in Kuhn's [15] sense.

IV. CONCEPTUAL FRAMEWORK AND RESEARCH PROPOSITIONS

Our conceptual framework integrates the typology with its antecedents and consequences. The AI capability level and the researcher's epistemic stance jointly determine the configuration adopted, which in turn shapes knowledge production outcomes and raises specific governance challenges. From this framework, we derive five research propositions:

P1: AI integration as an epistemic co-agent enhances the scope and efficiency of knowledge production relative to purely human-led processes, by expanding the hypothesis space through non-obvious pattern recognition and large-scale synthesis.

P2: The transition across configurations transforms the competencies required of researchers, shifting emphasis from data generation to critical evaluation and epistemic oversight, as AI progressively assumes execution tasks previously defining disciplinary expertise.

P3: Epistemic co-agency introduces new validity risks—including hallucination, algorithmic bias, and non-reproducibility—that require adapted governance mechanisms, because opacity and scale amplify error propagation across the research pipeline.

P4: Increased reliance on AI reshapes epistemic authority, potentially redistributing legitimacy from individual researchers to human-AI systems and destabilizing traditional attribution and peer-review frameworks.

P5: Organizational readiness—encompassing infrastructure, competencies, and institutional norms—moderates the effectiveness of human-AI epistemic collaboration, with outcomes bounded by the institution's capacity to absorb and validate AI-generated knowledge claims.

V. ORGANIZATIONAL AND GOVERNANCE IMPLICATIONS

A. Transforming Research Institutions

The emergence of AI as an epistemic co-agent requires research institutions to rethink their organizational architectures. The researcher's role evolves from solitary producer to orchestrator of human-AI epistemic systems, demanding new competencies: prompt engineering, critical evaluation of AI outputs, and epistemic

responsibility [14]. The concentration of advanced AI models in the hands of a few technology companies creates asymmetric access, raising equity concerns for under-resourced academic communities [7].

B. Epistemic Governance and Accountability

Governing AI as an epistemic co-agent requires mechanisms addressing three distinct risks. **Validity risks** concern the reliability of AI-generated knowledge ([2], [16]). **Legitimacy risks** concern the epistemic authority of AI-produced outputs [11]. **Dependency risks** concern the structural reliance on proprietary AI systems. The OECD [17] framework for anticipatory AI governance and the EU AI Act [5] provide relevant starting points for policy response.

VI. DISCUSSION

A. Theoretical Contributions

This paper makes three theoretical contributions. First, it conceptualizes AI as an epistemic co-agent, extending Dellermann et al.'s hybrid intelligence framework [4] into epistemic processes. Second, it proposes a three-configuration typology that operationalizes the human-AI relationship in knowledge production. Third, it derives five research propositions connecting AI capability levels, researcher roles, and governance requirements.

B. Managerial and Societal Contributions

For research managers, the typology provides a practical decision tool: choosing the appropriate human-AI configuration requires explicit analysis of epistemic stakes, AI system capability, and governance resources in place. For policymakers, the framework highlights the urgency of anticipatory governance: the gap between AI epistemic capability and institutional readiness is growing rapidly [8].

VII. LIMITATIONS AND FUTURE RESEARCH

This paper is conceptual in nature. Its propositions are derived from theoretical reasoning and illustrative empirical cases rather than systematic empirical testing. Future research should pursue empirical validation through case studies and surveys, longitudinal studies tracking how researcher roles evolve as AI capability increases, and comparative institutional analyses examining how governance frameworks mediate the risks of epistemic co-agency.

VIII. CONCLUSION

The question is no longer whether AI participates in knowledge production. The events of early 2026—Aletheia's autonomous research paper, the AI Co-scientist's validated biomedical hypotheses, the Genesis Mission's institutional deployment—confirm that it does. The question now is how to conceptualize, govern, and organize this participation responsibly.

This paper has proposed the concept of the epistemic co-agent as a theoretical anchor for that inquiry. By positioning AI as a genuine participant in knowledge construction and mapping the configurations through which this participation takes shape, we offer a framework for researchers, institutions, and policymakers navigating this transformation. This framework opens several research avenues. Empirical validation of the three-configuration typology could proceed through comparative case studies of AI-augmented research laboratories across disciplines. Longitudinal designs could track how configurations evolve as AI capability matures and as researchers develop epistemic literacy. Cross-institutional analyses could examine how governance frameworks mediate the validity, legitimacy, and dependency risks identified here. Finally, the emergence of fully agentic systems such as Aletheia [6] and AI Co-scientist [10] raises a provocative question: does autonomous machine discovery require a fourth configuration—Human as Auditor—in which epistemic oversight operates after rather than during knowledge production? Beyond the scientific domain, epistemic co-agency invites a broader societal reflection on how knowledge is produced, legitimized, and governed in AI-augmented societies—with profound implications for education, the democratization of science, and international scientific collaboration.

REFERENCES

- [1] R. Alvarado, "AI as an epistemic technology," *Science and Engineering Ethics*, vol. 29, no. 5, p. 32, 2023.
- [2] Z. Chen et al., "Research integrity in the era of artificial intelligence," *Medicine*, vol. 103, no. 29, 2024.
- [3] M. Coeckelbergh, "AI and epistemic agency: How AI influences belief revision," *Social Epistemology*, 2025.
- [4] D. Dellermann, P. Ebel, M. Söllner, and J. M. Leimeister, "Hybrid intelligence," *Business & Information Systems Engineering*, vol. 61, no. 5, pp. 637–643, 2019.
- [5] European Parliament, "Regulation (EU) 2024/1689—Artificial Intelligence Act," *Official Journal of the European Union*, 2024.
- [6] T. Feng et al., "Towards autonomous mathematics research," *arXiv preprint arXiv:2602.10177*, 2026.
- [7] Frontiers in Research Metrics and Analytics, "Open science falling behind in the era of artificial intelligence," 2025.
- [8] GESDA, "AI in 2026: From breakthrough to coordination," *Science Breakthrough Radar*, 2026.
- [9] Google Cloud, "Accelerating the Genesis Mission with Gemini for Government," Technical Report, Google, 2025.
- [10] Google Research, "Accelerating scientific breakthroughs with an AI co-scientist," Technical Report, Google DeepMind, 2025.
- [11] R. Hauswald, "Artificial epistemic authorities," *Social Epistemology*, vol. 39, no. 6, pp. 716–725, 2025.
- [12] P. Hemmer et al., "Complementarity in human-AI collaboration," *Journal of Information Technology*, 2025.
- [13] I. V. Kozlenkova et al., "Conceptual research: Multidisciplinary insights for marketing," *Journal of Marketing*, 2024.
- [14] R. Levy et al., "Epistemic responsibility: Toward a community standard for human-AI collaborations," *Frontiers in Psychology*, 2025.
- [15] T. S. Kuhn, *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press, 1962.
- [16] NIST, "Towards a standard for identifying and managing bias in AI," Special Publication 1270, 2022.
- [17] OECD, "Futures of global AI governance," OECD Publishing, 2024.
- [18] W. J. Orlikowski, "Sociomaterial practices: Exploring technology at work," *Organization Studies*, vol. 28, no. 9, pp. 1435–1448, 2007.
- [19] C. Rughinş et al., "AI at the knowledge gates," *Frontiers in Computer Science*, 2025.
- [20] Y. Xu et al., "Artificial intelligence: A powerful paradigm for scientific research," *The Innovation*, vol. 2, no. 4, 2021.
- [21] J. P. Cornelissen, "Editor's comments: Developing propositions, a process model, or a typology? Addressing the challenges of writing theory without a boilerplate," *Academy of Management Review*, vol. 42, no. 1, pp. 1–9, 2017.