

Rethinking AI's Role in Science: Toward a Theory of Epistemic Co-Agency and Its Implications for Research Organizations

Aya Daruich, Mounya Chahboune, Lhoussein Baouzil

LEG Research Unit, FSJES — Sultan Moulay Slimane University, Khouribga, Morocco

ayadaruich09@gmail.com; cmounya@gmail.com; lhbaouzil@gmail.com

Abstract— The question of how artificial intelligence is reshaping scientific knowledge production has moved, in a few short years, from theoretical speculation to practical urgency. Yet most organizational and management research still treats AI as a sophisticated tool — useful, even transformative, but ultimately passive in the epistemic sense. This paper argues that framing has become inadequate. Drawing on a theory-building conceptual approach, we develop the notion of AI as an epistemic co-agent: an active participant in generating, validating, and legitimating scientific knowledge alongside the human researcher. We organize this argument around a three-configuration typology — researcher as producer, collaborator, or supervisor — that evolves as AI autonomy rises. Five empirically testable propositions follow, and the paper ends by examining what this transition demands of institutions responsible for governing science.

Keywords— AI and epistemic agency; hybrid intelligence; knowledge production and governance; human-AI collaboration; sociomateriality; organizational practices; research integrity.

I. INTRODUCTION

In February 2026, a system called Aletheia, developed by Google DeepMind, published what may be the first fully autonomous mathematical research paper — calculating structure constants in arithmetic geometry without any human direction or intervention [6]. Days earlier, the same lab's AI Co-scientist had independently identified drug repurposing hypotheses whose validity matched findings that teams of human researchers had spent years establishing [10]. The following are not peculiarities of the eye. They are markers of something deeper, a change like the actors (or actors) who create scientific knowledge.

AI has been around for decades in research labs but has evolved in important ways. Consider the case in which an AI makes a hypothesis, which is then tested by the researcher and published in a paper. Who is that, then, who helped him? The straight answer is that there are no good frameworks for deciding. The conventions of authorship, accountability, and credit now seem to be on shaky ground.

There is a significant gap in the literature. AI as an efficiency tool has received a lot of attention in the field of organizational and management studies, where it is seen as a means of automating tasks, speeding up data processing, and aiding decision-making. Less is mentioned about AI's role in the epistemic process, which involves the production of knowledge, its verification, and social legitimization. In philosophy of science and social epistemology, some steps have been taken that are starting to tackle it, but not much has been done that relates to the questions of practice and institutions in terms of organization and governance.

This paper makes this attempt. The question we ask is: What are the consequences for how knowledge is produced and managed for the introduction of AI as an epistemic co-agent to scientific research processes and organizational practices? Theory building is a conceptual approach [13] that was carried out by conducting an integrative review of the literature of epistemology, information systems, organizational theory, and AI research. This is not an attempt to find a resolution for debates of an empirical nature — that is left for future work — but to create a theory that will be sufficient to the phenomenon in question.

The paper is organized as follows. Section II reviews four strands of literature: AI in scientific research, epistemic agency, hybrid intelligence, and organizational governance. In Section III, the epistemic co-agent concept is developed. The typology for the three configurations is presented in Section IV. The framework and propositions are given in Section V. Section VI takes up institutional and governance implications. Section VII ends with what's not resolved.

II. BACKGROUND AND LITERATURE

A. How AI's Role in Research Has Shifted

AI has been around for a while in scientific research, but usually as an assistant, for example, in running statistical scripts, collating bibliographic information, and processing the mechanical component of analysis. It's not that AI now does more of these things. Because of the nature of the tasks themselves, things have changed. Implementing such autonomous hypothesis generation, large-scale simulation, and drawing multi-disciplinary conclusions is, as described by Xu et al. [20], a paradigm shift: modern AI systems can come up with hypotheses without human intervention, conduct large-scale simulations, and now synthesize results across multiple disciplines, a scale individual researchers could not achieve.

The record of the ages is accumulating evidence to support this description. AlphaFold's breakthrough in solving the protein-folding problem, one that has been a challenge for structural biologists for half a century, is arguably the most cited: its results are increasingly being published in peer-reviewed journals as essential scientific contributions. Google DeepMind's AI Co-scientist proposes fresh ideas in biomedical research that have been confirmed by years of independent experiments [10]. The Genesis Mission's goal is to boost research productivity at institutional scale using agentic AI systems at U.S. national laboratories [9].

In about 10 years, changes have occurred that would have taken previous technological generations to realize: restructuring and redesigning the institutions that have formed around them. The capabilities of AI are significantly broader than what current oversight frameworks were designed for, and this is expanding. Chief among them was the human-centric model of knowledge production, which is no longer the only one present in research governance.

B. Epistemic Agency: What It Means and Why It Matters Here

In the broad sense, epistemic agency is the ability to contribute to the production of knowledge, to determine what is considered 'knowledge' and why, rather than merely storing or relaying knowledge [3]. There's little doubt that AI systems don't have anything remotely like this ability, but the discussion is no longer philosophical; it's impacting hiring practices, publication standards, and research funding.

Alvarado [1] believes that AI can be regarded as an epistemic technology, which is a technology of inquiry and a technology that can manipulate epistemic content in a way that is beyond information retrieval. While it is valuable, it may not be quite complete: it treats AI as a means of discovery rather than a means to generate the knowledge itself, namely, hypotheses and interpretations to be tested by researchers. Hauswald [11] further takes up the issue of legitimate knowledge claims in scientific communities that could be made by the output of AI and sets out to outline the standards that would need to be met for it to be legitimate.

Coeckelberg [3] moves the focus from knowledge of AI to its role in us, namely, about how much AI is integrated into practices that researchers use to create, test, and update beliefs. It's not actually the quality of the information that's the issue. The reason is that it changes the process itself, imperceptible and at a large level. It is within the context of norms of publication, peer review, and standards of attribution that these were developed in relation to a specific model of knowledge-making. That's a model that is being transformed by them.

C. Hybrid Intelligence and Human-AI Collaboration

Dellermann et al. [4] have introduced the notion of hybrid intelligence as a useful starting point in discussing knowledge work in human-AI teams. Their framework describes a sociotechnical configuration in which human and artificial intelligence can complement their capacities, in particular human contextual judgment, creativity, and normative reasoning, with AI's processing power, ability to recognize patterns in large volumes of data, and freedom from cognitive biases that can impact individual researchers. This claim is not

that AI takes the place of human intelligence, but rather that certain synergies result in a better performance than either AI or human intelligence individually.

Empirically, this sense that human-AI collaboration results in better performance is substantiated by Hemmer et al. [12], who show that, in fact, in complex decision-making tasks, it is just within the strategic management of capability asymmetries that human-AI collaboration results in superior performance. Rughinis et al. [19] take this argument further and apply it to the particular realm of scientific publishing, seeing AI as functioning on the knowledge gates: the filters that make claims epistemic within a discipline. A related thread is picked up by Levy et al. [14], who contend that there are community-level standards for epistemic responsibility in human-AI collaborations.

D. Organizational and Governance Concerns

Governance considerations of the integration of AI in science are starting to be systemically addressed, but are less developed than the growth of capabilities. The issue of reproducibility in AI-enhanced science is a growing concern, as noted in *Frontiers in Research Metrics* [7], due to dynamic changes in LLM versions and a lack of transparency in data and training methods. A more severe danger is the ability of AI to produce believable-sounding falsehoods, which can be used to support the kinds of research misconduct that are more difficult to detect via the traditional peer review process, as Chen et al. [2] explain.

Implicitly, NIST [16] has identified algorithmic bias as an issue that is not confined to any single output; it filters through into the “knowledge landscape” that is then used in the next steps of research. Artificial Intelligence is at the forefront of the scientific agenda for this decade, as institutions are responding to what they can, rather than what they can foresee, according to GESDA's 2026 Science Breakthrough Radar [8]. Both OECD [17] and EU AI Act [5] advocate a more proactive approach, but not specifically in the context of scientific research.

III. AI AS AN EPISTEMIC CO-AGENT

A. Three Positions AI Can Occupy in Research

Consider icons or other symbols that are common in their research, such as a reference management program, a statistics program, or a program to clean data. These are the kinds of things we have in mind when we talk about AI-as-tool: useful without any knowledge, epistemically inert. One step further, AI-as-assistant: the system recommends papers, detects something out of the ordinary in the dataset, and condenses a set of papers into a readable summary. The researcher is still driving, but the process has begun to take shape. The third position, epistemic co-agent, is of a different kind. If the researcher thinks of something that AI didn't, or if it discovers a pattern that changes the entire framework of how the researcher thinks about the study, the results are not the researcher's. It was from another place as well.

This third position is important to differentiate from related concepts carefully. Not what kind of knowledge is produced and whose, but rather, whether or not hybrid intelligence [4] is superior to human intelligence. Epistemic technologies [1] consider AI as an instrument for inquiries, but do not make it a participant. While helpful in understanding entanglement, the concept of sociomateriality [18] is not very informative with respect to how scientific validity is created in entangled systems. Epistemic co-agency is a more specific concept: It refers to the active production, testing, and validation of claims by AI that become part of the scientific record as knowledge.

B. Four Dimensions That Define the Degree of Co-Agency

AI systems are not always co-agents and can be different in different contexts. We propose 4 dimensions to capture this variation. Cognitive contribution is about whether AI is creating new epistemic content, such as hypotheses, interpretive moves, conceptual links, etc., or simply working with existing content. The degree of functional autonomy of AI is how autonomous it can be from the researcher's direction. Interaction quality reflects the critical depth of the human-AI interaction. And influence on validation – are the answers we as a research community prioritize becoming what AI outputs – not what we write, but what we prioritize for acceptance?

IV. THREE CONFIGURATIONS OF HUMAN-AI RESEARCH

A. Two Axes, Three Positions

We have organized the typology along two axes according to the hybrid intelligence framework of Dellermann et al. [4] with the extension of this framework to the epistemic domain. The first is functional autonomy — from narrow task-specific systems to frontier generative models, to autonomous fully agentic systems that initiate and close research tasks on their own. The second is the epistemic stance of the researcher, which does not stand alone: the more autonomous the AI, the more it becomes a partner, co-creator, and/or supervisor.

B. Configuration 1 — Human as Producer

The first configuration is one in which the researcher is fully responsible for the epistemic initiative. AI is not a tool that does the thinking work or the evaluation work; it serves as a tool or assistant to run, retrieve, and organize the literature, to collect the data, whatever it is that is computational. Knowledge production is from hypothesis to conclusion – AI is about execution, not reasoning. It is the predominant way AI will be used in research for the early 2020s. The challenges it faces in governance are similar to those found in conventional research ethics: correct attribution and data management, and methodological transparency.

C. Configuration 2 — Human as Collaborator

With the second setup, the division of work between man and machine becomes progressively blurred. Researcher + AI, directions set by researcher, options provided by AI, assessed + redirected by researcher, refined by AI. By the time a finding is up for writing, it's impossible to separate the parts of the work contributed by each party. An example of this is the case of Google AI Co-scientist [10]: scientists formulate general questions, AI suggests and filters particular ideas, and the studies that would have been created without the assistance of AI would not be the same. It is a transparent world with respect to what AI offered, attribution criteria, and review processes sensitive to the failure modes of AI.

D. Configuration 3 — Human as Supervisor

The third configuration has AI taking a dominant epistemic role, creating, checking, and formalizing knowledge outputs with little human input. The role of the researcher ranges from knowledge production to post-hoc evaluation and institutional validation. The most clear-cut example of this today is Aletheia's mathematical work [6], which is independent. Three epistemological problems arise in a natural way in this level of AI autonomy: circular validation of the production and validation of knowledge by the same system; diffuse epistemic responsibility; and the structural inability of fully autonomous AI to have the paradigmatic change that Kuhn [15] considers to be the motor of scientific advances.

TABLE I
TYPOLOGY OF HUMAN-AI EPISTEMIC CONFIGURATIONS

Dimension	Config. 1 — Human as Producer	Config. 2 — Human as Collaborator	Config. 3 — Human as Supervisor	Governance Risk
AI Autonomy	Low — narrow AI systems	Medium — generative AI models	High — fully agentic AI	—
Researcher Role	Holds full epistemic initiative; AI executes defined tasks	Co-constructs knowledge through iterative dialogue with AI	Evaluates and validates AI-generated outputs post hoc	Diffuse accountability
AI Contribution	Task execution only; no ideation or hypothesis generation	Hypothesis generation, pattern identification, interpretive support	Primary knowledge production and self-verification	Circular validation risk

Dimension	Config. 1 — Human as Producer	Config. 2 — Human as Collaborator	Config. 3 — Human as Supervisor	Governance Risk
Illustrative Example	Statistical software, automated bibliometrics, data cleaning tools	Google DeepMind AI Co-scientist [10]	Aletheia autonomous mathematical research paper [6]	Attribution and legitimacy uncertainty
Governance Mechanism	Standard research ethics, proper attribution, data management norms	Transparency protocols, contribution disclosure, critical oversight	Mandatory human audit, strict reproducibility requirements	Institutional AI governance frameworks

Source: Authors, adapted from Dellermann et al. [4]

V. FRAMEWORK AND RESEARCH PROPOSITIONS

A. The Integrated Framework

The conceptual framework that we propose combines the three configurations and their determinants and consequences (Fig. 1). This combination of AI capability level (narrow to fully agentic) and epistemic stance helps to determine the configuration for a particular institutional context. The configuration, in turn, determines knowledge production outcomes – in scope and efficiency, in terms of risks of invalidity, and in terms of governance requirements. These relationships are moderated by two moderating variables, organizational readiness and governance maturity. They put together the reasons for the different knowledge quality results in institutions with the same AI capabilities.

There is also a dynamic aspect to the framework, as the capabilities of AI evolve, configurations that are currently outstanding will become increasingly common. Any research institutions that have just created governance frameworks for only Configuration 1 will discover that they are not ready for the shift of the center of gravity.

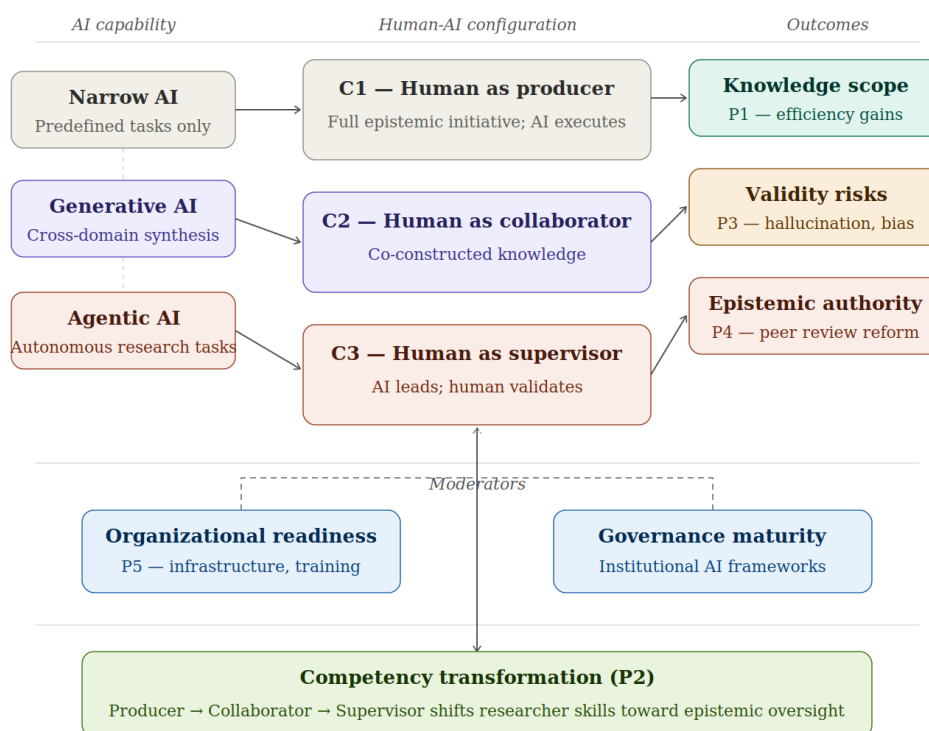


Fig. 1 — Conceptual framework of human-AI epistemic co-agency. Source: Authors.

Fig. 1 — Conceptual framework of human-AI epistemic co-agency

Source: Authors

B. Five Propositions

P1 deals with the efficiency and breadth of knowledge. In the case of AI acting as an epistemic co-agent, namely in Configuration 2, it increases the hypothesis space by non-obvious pattern recognition of large and heterogeneous data. The breadth and speed advantages of knowledge production imply hypotheses that a researcher would not have achieved on his/her own, not because of the lack of effort but because of the structural invisibility of the respective patterns to the naked eye of a human being.

P2 has to do with competency change. With the change of the role of the researcher as a producer to collaborator and supervisor, the competencies that form the disciplinary expertise are recomposed structurally. Activities that were traditionally at the core of research identity, such as experimental design, statistical analysis, and literature synthesis, are increasingly being taken over by AI, and new skills, such as critical assessment of AI results, epistemic responsibility, and AI literacy in a research-specific sense, are increasingly valuable. Re-designing of doctoral training, faculty evaluation and research assessment systems will be necessary.

P3 distinguishes three types of validity risk that epistemic co-agency brings and that are ill-suited to be identified by current peer review. Hallucination - the assured generation of plausible, but factually incorrect material - is a qualitatively new form of error, in that it is not obviously incorrect productions that are similar to correct productions. Algorithm bias works are trained in such a way that the reviewers who do not know about these technical aspects cannot see it. The failure to reproduce the versions of a model puts the cumulative progress of science at risk.

P4 deals with the issue of the redistribution of epistemic authority. The traditional sources of scientific authority, including knowledge as demonstrated in published work, institutional membership, peer-reviewing, etc., are somewhat substituted by AI systems that have more epistemic roles. The resultant redistribution will put epistemic power in the hands of institutions and organizations that regulate access to frontier AI systems, and the consequences of this on global research equity are yet to be adequately tackled by policy frameworks.

P5 discusses the moderator of organizational preparedness. The relationship between the capacity of AI and the quality of knowledge is not direct but through the institutional conditions. Preparedness is multidimensional - it needs to have the computational infrastructure, trained human capital, aligned incentive structures, and operating governance structures - and lack of any of the dimensions can weaken the others. The policies and interventions should also be multidimensional; investments in the infrastructure should be done without giving the same consideration to training and governance; otherwise, the results will not be optimal.

TABLE II
RESEARCH PROPOSITIONS AND KEY CONCEPTS

Prop.	Statement	Key Concepts
P1	When AI is employed as an epistemic co-agent, by non-obvious pattern recognition and massive hypothesis generation, which is not possible to be produced by individual researchers, the scope and power of knowledge generation is increased.	Hypothesis space; knowledge efficiency; AI-augmented discovery
P2	As the relations between humans and AI transform into producer-collaborator and supervisor, the competencies of the researcher are restructured - not to produce data, but to critically assess it, have epistemic control, and be AI literate.	Skills transformation; epistemic literacy; competency redesign
P3	Systemic validity risks, such as AI hallucination, propagation of algorithmic bias, and non-reproducibility, are introduced by epistemic co-agency, and are not well-addressed by current peer-review mechanisms.	AI validity; research integrity; hallucination; reproducibility
P4	The additional entrenchment of AI shifts the balance of epistemic power, in part, substituting the individual researcher's knowledge and disrupting the traditional peer-review system.	Attribution; epistemic authority; peer-review reform
P5	One of the primary mediators of the quality of epistemic collaboration is organizational preparedness: institutions with poor infrastructure, trained personnel, and governance framework will yield inferior outcomes with the same amount of AI capabilities in a systematic way.	Institutional capacity; AI readiness; governance maturity

Source: Authors

VI. WHAT THIS MEANS FOR INSTITUTIONS

A. The Organizational Challenge

As long as the argument made in this paper holds, the problem of adaptation of research institutions is more basic than most of the AI policy debates of today are making it out to be. It is not only how to regulate the use of AI in the existing research paradigms, but how to redesign the research paradigms to a knowledge production that is actually different than what they were intended to achieve. The task of the researcher is becoming that of a single knowledge producer to that of an organizer of human-AI epistemic systems - and the task demands new competencies, new evaluation criteria, and new frameworks.

The competency dimension is particularly sharp in doctoral training, where in most disciplines, the training is organized in terms of Configuration 1 competencies. The more common Configuration 2 and 3 environments are the larger the disjuncture between what is generated by programs and what research practice needs will be. Structural equity is also an issue: the concentration of the frontier AI capabilities in a few technology firms generates asymmetries in research capacity with no apparent historical basis. When AI co-agency really widens the scope and efficiency of knowledge production, these asymmetries will increase the international scientific capacity gap unless action is taken, e.g., by investing in open-source AI infrastructure [7].

B. Governance Mechanisms

To control AI as an epistemic co-agent, there is a need to have mechanisms that will deal with three categories of risks. Threats of validity, hallucination, algorithmic bias, and non-reproducibility demand

technical governance: compulsory disclosure of AI tools and model versions, reproducibility criteria defining AI system configurations, and the creation of AI-specific peer review skills in the disciplinary communities.

The legitimacy risks need to be regulated by the institutions: new authorship policies that directly address AI contributions, contribution taxonomies that help to understand the role of humans and AI in knowledge creation, and accountability systems that help to understand the role of the researcher in AI-enhanced outputs [11]. Risks of dependency must be addressed strategically: investing in open-source AI infrastructure, regulation to avoid vendor lock-in, and standards and governance that require research-grade AI systems to have stable, documented versions. The initial points of departure are found in OECD [17], and the EU AI Act [5], but none of them was specifically designed to be used in research. The point is that the governance in both cases should not be equal, but proportional, i.e., the governance of Configuration 3 should be stricter than that of Configuration 1.

VII. CONCLUSION

The use of artificial intelligence in scientific research is no longer a future eventuality - it is a reality, and it is becoming increasingly popular with repercussions that the existing institutional frameworks are only beginning to think about. The concept of AI as an epistemic co-agent, three-configuration typology, and five research propositions that connect the degrees of AI capability, role of researchers, and the governance needs are the theoretical contribution of the paper.

Science as a professional practice is at stake much more than science itself. Scientific knowledge production is one of the key processes through which truth claims are produced, evaluated, and empowered in most of the modern societies. The implementation of AI in this process has an implication on how societies resolve conflicts, distribute resources, and negotiate collective issues. The question of who/what a legitimate producer of knowledge is is not only an academic one: it affects the social infrastructure in which evidence is turned into policy, education, and discourse.

There are several weaknesses of the current work. The conceptual basis of the typology is not empirically tested; it is still to be discovered whether the three configurations can be differentiated in the research practice and whether the propositions can be used in other fields and in the institutional context. The consequences of governance are introduced on the premise of first principles and new policy frameworks, and their efficacy will have to be closely evaluated when they are implemented by the institutions.

Such restrictions provide the orientation of future studies. The case studies of comparison of various disciplines can be used to establish whether the configurations can be empirically distinguished. The longitudinal studies would be in a position to trace the evolution of the role of researchers and the organization of governance with regard to the evolution of AI capability. Cross-institutional analyses of the mediation of the governance structures could be used to assess the validity, legitimacy, and dependency risks identified in the propositions. With the increase in the number of fully agentic systems, a fourth configuration, Human as Auditor, might need to be theorized. What seems to be self-evident is that the change that is underway is not so much a change in technology as it is a change in institutions and civilizations: it is about the social systems within which human communities determine what knowledge is, who is a producer of knowledge, and on what terms.

REFERENCES

- [1] R. Alvarado, "AI as an epistemic technology," *Science and Engineering Ethics*, vol. 29, no. 5, p. 32, 2023.
- [2] Z. Chen, Y. Liu, H. Wang, and J. Li, "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*, vol. 39, no. 2, pp. 101-115, 2025.
- [4] D. Dellermann, P. Ebel, M. Sollner, 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, L 1689, 2024.
- [6] T. Feng, Z. Lin, Y. Wang, and D. Gao, "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: Challenges and recommendations," Frontiers, vol. 9, 2025.
- [8] GESDA, "AI in 2026: From breakthrough to coordination," Science Breakthrough Radar, Geneva Science and Diplomacy Anticipator, 2026.
- [9] Google Cloud, "Accelerating the Genesis Mission with Gemini for Government," Technical Report, Google, Mountain View, CA, 2025.
- [10] Google Research, "Accelerating scientific breakthroughs with an AI co-scientist," Technical Report, Google DeepMind, London, 2025.
- [11] R. Hauswald, "Artificial epistemic authorities," Social Epistemology, vol. 39, no. 6, pp. 716-725, 2025.
- [12] P. Hemmer, M. Schemmer, N. Kuhl, M. Benz, and G. Satzger, "Complementarity in human-AI collaboration: Concept, drivers, and practical implications," Journal of Information Technology, vol. 40, no. 1, pp. 22-41, 2025.
- [13] I. V. Kozlenkova, R. W. Palmatier, E. Fang, B. Xiao, and M. Houston, "Conceptual research: Multidisciplinary insights for marketing," Journal of Marketing, vol. 88, no. 3, pp. 1-22, 2024.
- [14] R. Levy, A. Cohen, and T. Granot, "Epistemic responsibility: Toward a community standard for human-AI collaborations," Frontiers in Psychology, vol. 16, 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 artificial intelligence," Special Publication 1270, National Institute of Standards and Technology, Gaithersburg, MD, 2022.
- [17] OECD, "Futures of global AI governance," OECD Publishing, Paris, 2024.
- [18] W. J. Orlikowski, "Sociomaterial practices: Exploring technology at work," Organization Studies, vol. 28, no. 9, pp. 1435-1448, 2007.
- [19] C. Rughinis, S. Rughinis, and D. Vulpe, "AI at the knowledge gates: Epistemic power and machine learning in scientific publishing," Frontiers in Computer Science, vol. 7, 2025.
- [20] Y. Xu, X. Liu, X. Cao, C. Huang, E. Liu, and S. Qian, "Artificial intelligence: A powerful paradigm for scientific research," The Innovation, vol. 2, no. 4, p. 100179, 2021.
- [21] J. P. Cornelissen, "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.