# The Fifth Paradigm: Scientific Discovery in the Age of Autonomous Artificial Intelligence


### Abstract

Science is undergoing a paradigmatic transformation, transcending the era of data-intensive science (the Fourth Paradigm) to inaugurate a Fifth Paradigm, driven by Artificial Intelligence (AI). This article traces the evolution from the response to the "data deluge," conceptualized by Jim Gray, to the consolidation of a global data infrastructure that has become the foundation for the AI revolution. We analyze how AI architectures—notably Transformers and diffusion models—are redefining discovery in domains such as genomics, materials science, and the social sciences, shifting from data analysis to the autonomous generation of hypotheses. This transition culminates in the vision of the "robot scientist," which automates the complete scientific cycle. However, this new paradigm engenders profound crises. The epistemological crisis, centered on the "opacity" of AI models, challenges the concepts of scientific justification, explanation, and reproducibility. Simultaneously, an integrity crisis emerges, with the proliferation of AI-generated errors and fraud, exposing vulnerabilities in the academic publishing ecosystem. We conclude that the future of science lies not in the replacement of the human, but in a cognitive symbiosis. The role of the scientist evolves into that of a curator of questions, an ethical supervisor, and a critical partner to AI, orchestrating discovery through methodologies like Human-in-the-Loop (HITL) to ensure that AI's computational power augments, rather than supplants, the human quest for knowledge.


## 1. Introduction: From the Data Deluge to the Genesis of a New Paradigm

The history of science can be understood as a succession of paradigms, each defined by its methodologies, tools, and fundamental assumptions. The first three paradigms form the basis of modern scientific practice. The first, an empirical paradigm dating back millennia, focused on the description of natural phenomena.[1] In recent centuries, the second, theoretical paradigm emerged, using models and generalizations to explain observations, as exemplified by Newton's laws and Maxwell's equations.[3] In the last decades of the 20th century, the advent of high-performance computing gave rise to the third, computational paradigm, which allowed for the simulation of complex phenomena whose theoretical models were analytically intractable.[5]

However, at the dawn of the 21st century, science faced a crisis of a different nature. Increasingly sophisticated instruments—from sensors and genome sequencers to particle colliders and digital telescopes—along with supercomputer simulations, generated an unprecedented volume, variety, and velocity of data.[2] This "data deluge" was not just a quantitative challenge, but a methodological crisis.[2] As computing pioneer Jim Gray observed, scientists found their data in "digital shoeboxes," overwhelmed with information and with tools, like spreadsheets, that were rapidly becoming obsolete.[4] The bottleneck was no longer data generation, but its management, analysis, and interpretation.

In response to this crisis, the Fourth Paradigm emerged: data-intensive science or eScience.[1] Proposed by Gray and his collaborators, this new paradigm was not a mere extension of computational science, but a fundamentally new approach that unified theory, experiment, and simulation through data.[4] Its core methodology is based on three essential activities: **capture, curation, and analysis**.[7] Capture refers to the collection of data from various sources. Analysis uses statistical and modeling tools to extract knowledge. Crucially, curation—the organization, annotation, and preservation of data with explicit schemas and metadata—was identified as the pillar to ensure the longevity, interoperability, and reusability of data, preventing its interpretation from being trapped in specific software programs.[7]

Jim Gray used the metaphor of the "data iceberg" to illustrate that the published scientific literature represents only the visible tip of a vast volume of collected data that remains uncurated, unanalyzed, and unpublished systematically.[4] The goal of the Fourth Paradigm was, therefore, to make this submerged mass of data a living, accessible, and permanently available resource for the scientific community.[7] This vision implied a fundamental shift in the valuation of scientific products: raw data, derived data, and the software used to analyze them should be considered first-class objects, as important as the final research paper.[9] In doing so, the Fourth Paradigm not only resolved the data deluge crisis but, without fully foreseeing it, laid the cultural and technical groundwork for the next scientific revolution.

## 2. The Consolidation of the Fourth Paradigm: Infrastructure for Data-Driven Science

The vision of the Fourth Paradigm did not remain an abstract concept; it materialized in the construction of a robust global cyberinfrastructure, which became the necessary condition for the emergence of the subsequent AI paradigm. This consolidation occurred through the development of large-scale data repositories, the adoption of cloud computing, and the formalization of policies and principles governing data sharing.

One of the most emblematic examples of this infrastructure is the *Sloan Digital Sky Survey (SDSS)*, an astronomical project that not only collected terabytes of imaging and spectroscopy data but made them publicly accessible through the *SkyServer*, a web interface that allowed thousands of scientists to make new discoveries far beyond the scope of the original team.[2] In biology, the creation of data commons like the

*NCI Genomic Data Commons* and the *BloodPAC Data Commons* transformed cancer and liquid biopsy research, allowing massive genomic datasets to be shared and reanalyzed in new contexts to generate and test new hypotheses.[3] Infrastructures like JASMIN in the UK exemplify the integration of data storage, supercomputing, and private cloud to serve the environmental sciences community.[2]

The rise of cloud computing was a crucial technological catalyst, solving one of the biggest challenges of data-intensive science: the movement of massive datasets. By co-locating data and computational power, the cloud made it more efficient to "move the query to the data" rather than the other way around, a fundamental logic shift for petabyte-scale analysis.[4]

Parallel to technological development, the consolidation of the Fourth Paradigm was cemented by a framework of policies and principles. Funding agencies, such as the *National Science Foundation* (NSF) and the *National Institutes of Health* (NIH) in the US, began requiring Data Management Plans in all research proposals, formalizing data curation as an integral part of the scientific process.[9] The NIH Public Access Policy, which requires the deposit of peer-reviewed manuscripts in the PubMed Central digital archive, ensured that scientific literature remained intrinsically linked to the data supporting it.[2] The pinnacle of this formalization was the **FAIR Principles (Findable, Accessible, Interoperable, Reusable)**, published in 2016, which established a global standard for scientific data management, with a particular emphasis on interoperability and machine-processability.[5]

The success of this endeavor is evident. A decade after the publication of the seminal book "The Fourth Paradigm," many of its predictions have become reality. The *National Oceanic and Atmospheric Administration* (NOAA) collects more than 20 terabytes of data daily, and digital repositories like Dryad and Zenodo have become standard for storing research data.[9] Genomics, in particular, has become the exemplary field of data-intensive science, where sequencing a single genome can generate terabytes of information, requiring highly specialized laboratory information management systems (LIMS).[10]

However, the true strength of the Fourth Paradigm lay not just in the volume of data, but in the creation of an interconnected ecosystem. The NCBI's Entrez system, for example, does not just store data; it weaves a network of connections between databases of nucleotides, proteins, 3D structures, and the PubMed literature, allowing researchers—and, crucially, algorithms—to navigate different layers of biological information in an integrated way.[2] It was this infrastructure of interconnected, curated, and computationally accessible data that inadvertently provided the "fuel" and "engine" for the AI revolution. The vast genomic datasets, astronomical images, and digitized scientific texts became the indispensable training grounds for the deep learning models that would define the next scientific paradigm.

## 3. The AI Rupture: The Transition to a Fifth Paradigm

While the Fourth Paradigm was consolidating, a new technological force was emerging, a force that represented not a linear continuation, but a qualitative rupture. The rise of Artificial Intelligence (AI), especially deep learning and generative models, was not just an improvement in data analysis tools; it introduced a new capability into the scientific process: the automation of cognitive work. This development marks the transition to a Fifth Paradigm of scientific discovery.

It is telling that the 10-year retrospective analysis of the book "The Fourth Paradigm" highlights the "AI and Deep Learning Revolution" as one of the most significant developments of the decade, but one that was *not strongly identified* in the original

2009 work.[9] This indicates that AI was not seen as a natural evolution of data-intensive science, but as an exogenous and disruptive force that fundamentally altered the landscape.

The fundamental distinction between the Fourth and Fifth Paradigms lies in the shift of focus from **data analysis** to **autonomous hypothesis generation**.[12] The Fourth Paradigm provided the tools for scientists to explore vast datasets in search of patterns to validate or refute their own hypotheses.[7] The Fifth Paradigm, on the other hand, introduces systems that can generate plausible and testable hypotheses directly from the data, automating a step of the scientific method that was considered a bastion of human creativity and intuition.[14] AI is no longer just finding the needle in the haystack; it is starting to design new and better needles.

This shift redefines the role of AI in science, from a tool to a collaborator. The narrative evolves from AI as a powerful statistical analyzer to AI as a **"co-scientist"** or research partner.[14] The symbiotic collaboration between human and artificial intelligence is now seen as the engine of future scientific discovery.[17]

Thus, we can define this new paradigm. If the Fourth Paradigm is *data-intensive* science, the Fifth Paradigm is *data-intelligence-intensive research* [19] or, more directly,

*AI-driven science*.[12] This new era not only uses large-scale data but employs AI systems to integrate theory, experiment, simulation, and analysis in unprecedented ways, with AI acting as the connective tissue linking all phases of the research cycle.[20] The methodology ceases to be primarily exploratory and data-oriented to become hypothesis-generating and AI-driven.[20]

The most profound consequence of this transition is the shift in the *locus* of scientific creativity. A study by the *National Bureau of Economic Research* (NBER) revealed that an AI tool in a materials science lab automated 57% of "idea generation" tasks, shifting the scientists' work to the *evaluation* of hypotheses generated by the machine.[13] This inversion of the traditional workflow demonstrates that AI is not just accelerating science, but reconfiguring the very nature of creative work. If hypothesis generation—the inductive or abductive leap that defines discovery—can be partially or fully automated, it challenges long-standing philosophical notions about human intuition and ingenuity as the sole drivers of scientific progress, raising fundamental questions about what it means to "discover" in a world where machines can participate in that act.

## 4. The New Machines of Discovery: Foundational Architectures and Models

The Fifth Paradigm is driven by a new class of computational tools whose sophistication and generative capacity far surpass those of the previous era. Three AI architectures, in particular, form the core of the new discovery machines: the Transformer architecture, generative diffusion models, and large language models (LLMs). These technologies are not just more powerful; they operate under a unifying paradigm of treating diverse types of scientific data—from amino acid sequences to molecular structures and text—as "languages" to be learned and generated.

### 4.1. Transformer Architecture: Deciphering the Language of Nature

Originally developed for natural language processing, the Transformer architecture has proven to be a surprisingly universal tool for science.[21] Its power lies in the **self-attention** mechanism, which allows the model to weigh the importance of different parts of an input sequence, capturing long-range dependencies that were a challenge for previous recurrent architectures.[23] By treating scientific data as sequences—a chain of amino acids, a linear representation of a molecule (SMILES), or a set of temporal observations—Transformers can "read" the language of nature.

The most prominent case study is DeepMind's **AlphaFold**, which applied a Transformer-based architecture to solve the decades-old problem of protein folding.[26] By treating the amino acid sequence as a language, AlphaFold learned to predict the protein's three-dimensional structure with a precision comparable to experimental methods, unleashing a revolution in structural biology.[28] Subsequent research has massively expanded the use of Transformers to a vast range of problems in proteomics, including function prediction, protein-protein interaction analysis, and new drug discovery.[21]

In chemistry, Transformers are applied to tasks such as retrosynthesis planning (figuring out how to synthesize a molecule) and exploring the vast chemical space for new compounds.[21] Ambitious projects like Microsoft Research's **NatureLM** aim to create unified foundation models that use a single Transformer architecture to perform tasks in disparate scientific domains, such as molecules, proteins, DNA, and materials, treating them all as sequence modeling problems.[30] This convergence suggests that the "language" of discovery in the Fifth Paradigm is largely that of sequence modeling and its underlying probability distributions.

### 4.2. Generative Diffusion Models: Building Matter from Noise

While Transformers excel with sequential data, generative diffusion models have become the dominant technology for generating complex, high-dimensional data, such as 3D molecular structures. Inspired by non-equilibrium thermodynamics, these models work through a two-step process: first, they learn to systematically destroy the structure of training data by adding Gaussian noise step-by-step; then, they learn to reverse this process, generating new, realistic data from pure noise.[31] Their ability to model complex probability distributions has surpassed other generative models, such as Generative Adversarial Networks (GANs), in many tasks.[31]

In materials science and chemistry, diffusion models are driving a new era of *de novo* design.[35] They are used to generate 3D molecular structures with desired physical or chemical properties, directly addressing the challenge of inverse design.[37] Recent research (2024-2025) already demonstrates specialized diffusion models for discovering surface structures in materials [38] and hybrid models that combine the continuous generation of diffusion with the discrete structure of autoregressive models to improve the speed and quality of molecule generation.[37]

The power of these models for scientific discovery lies in their ability to transform the search for new materials into a conditional sampling problem.[34] Instead of blindly searching the vast chemical space, scientists can "ask" the model to generate structures that maximize a certain property (e.g., binding affinity to a drug target or thermal stability). This capacity for creative extrapolation—generating genuinely new artifacts that were not in the training data—is what distinguishes "discovery" from mere "analysis" and is a pillar of the Fifth Paradigm.

### 4.3. Large Language Models (LLMs): Simulating and Analyzing the Social World

Large Language Models (LLMs), such as the GPT family, are a specialized application of the Transformer architecture, trained on massive-scale text corpora. Their main strength lies in their ability to process unstructured text, infer meaning from context, and perform complex reasoning and language generation tasks.[40]

In computational social sciences, LLMs are transforming classic research methods.[42] They are employed in:

- **Content Analysis:** To classify texts according to complex categories (topics, frames, sentiment) with little or no need for manually labeled data (*zero-shot* or *few-shot classification*), a task that previously required intensive human labor.[43]
- **Survey Research:** To simulate the responses of different subpopulations to survey questions. By fine-tuning LLMs to reflect specific demographic or ideological profiles, researchers can model public opinion at scale and explore hypothetical scenarios without the costs and time associated with traditional surveys.[42]
- **Experimental Studies:** To generate dynamic and personalized experimental stimuli, and to simulate interactions between agents in complex social models, allowing the study of phenomena like polarization or information dissemination in controlled environments.[42]

Despite their potential, the application of LLMs in the social sciences faces unique challenges. The lack of deep domain knowledge can lead to superficial analyses, and the risk of "hallucinations"—the generation of factually incorrect but plausible information—is particularly dangerous in research contexts.[40] Furthermore, biases present in the vast internet training data can be reproduced and amplified in the results, and ensuring the reproducibility of results with models that are constantly being updated is a significant methodological challenge.[43]

## 5. The Autonomous Frontier: The Advent of the Robot Scientist

The most advanced and transformative manifestation of the Fifth Paradigm is the convergence of cognitive artificial intelligence with physical automation, giving rise to the concept of the **robot scientist**. This is not just an AI system that analyzes data, but an integrated platform that aims to automate the entire cycle of scientific discovery, from hypothesis formulation to experimental validation and dissemination of results.[49]

A robot scientist is defined as a system that employs AI to generate hypotheses from a computational model of a domain, designs experiments to test these hypotheses, physically executes these experiments using laboratory robotics, analyzes the resulting data, and, crucially, uses these results to refine its hypotheses and initiate the next discovery cycle. This process, known as *closed-loop learning*, represents the end-to-end automation of the scientific method.[49]

The realization of this vision depends on the synergy between two key technologies: **agentic AI** and **embodied robotics**.[51] Agentic AI, often based on LLMs, takes on the cognitive tasks: reviewing literature to identify knowledge gaps, generating hypotheses, planning experimental protocols, and even writing scientific manuscripts.[51] Embodied robotics, in turn, translates these digital plans into actions in the physical world, manipulating samples, operating instruments, and collecting data in autonomous laboratories.[53]

The methodologies for autonomous hypothesis generation, the cognitive heart of the robot scientist, are already being actively developed. As detailed in recent reviews, these include [50]:

1. **Literature-Based Knowledge Graphs:** AI synthesizes the vast scientific literature into a knowledge graph, analyzing it to propose undiscovered links between concepts, genes, or compounds.
2. **Symbolic Regression:** AI searches for the underlying mathematical equations that best describe a dataset, formulating physical or biological laws in a symbolic form.
3. **Human-AI Collaboration:** Systems where AI generates a large number of candidate hypotheses and the human scientist uses their intuition and domain knowledge to select the most promising ones for testing.
4. **Generative Creative Thinking:** The use of generative AI to explore the hypothesis space in ways that humans, with their cognitive biases, might not consider.

The potential impact of this total automation is immense. It promises not only to drastically accelerate the pace of discovery but also to increase the precision and reproducibility of experiments by minimizing human error.[49] However, the rise of the robot scientist fundamentally changes the nature of experimentation itself. Science ceases to be a slow, deliberate process guided by human intuition and becomes a high-throughput optimization process, where thousands of hypotheses can be generated and tested in parallel or in rapid succession. Discovery transforms from an act of "insight" to an act of "optimized search."

The ultimate vision goes beyond specialized robots in narrow domains. The concept of an **"Autonomous Generalist Scientist" (AGS)** suggests a system capable of integrating knowledge across diverse scientific disciplines.[51] The ability to perform interdisciplinary synthesis—one of the greatest challenges of modern science—may be the most revolutionary contribution of the robot scientist, enabling breakthroughs in complex problems that lie at the intersection of fields like biology, physics, and materials science.

## 6. The Epistemological Crisis: Justification, Opacity, and the Nature of Scientific Knowledge

The rise of the Fifth Paradigm, while promising, triggers a profound epistemological crisis, challenging the very foundations of how scientific knowledge is generated, justified, and understood. At the heart of this crisis is the problem of **opacity** in AI models, particularly those of deep learning.[55]

These models are often described as "black-box models" because, although they can produce highly accurate results, their internal processes are of such complexity that they become unintelligible to humans.[57] Epistemic opacity refers to the impossibility for a human agent to know all the elements and logical rules that govern the transformation of inputs into outputs within the neural network.[56] This is not just a matter of trade secrets, but an inherent characteristic of their architecture and the way they learn autonomously from data.[57]

This opacity poses a direct challenge to the pillars of science. The **justification** of a scientific claim traditionally requires an explicit and verifiable chain of reasoning. If the justification is based on the output of an opaque model, this chain is broken, making it difficult to assess the validity of the claim.[56] Similarly, scientific **explanation** goes beyond mere prediction; it seeks a causal understanding of phenomena. Opaque models can predict accurately but fail to provide comprehensible explanations, shifting the focus of AI philosophy from "can machines think?" to the more pressing question of "how do machines think?".[60]

**Reproducibility**, another scientific pillar, is also threatened, as the lack of transparency and insufficient documentation make it almost impossible for other researchers to replicate results obtained with complex AI tools.[55]

A philosophical response to this dilemma proposes a distinction between the **"context of discovery"** and the **"context of justification"**.[56] In this view, opaque AI can be legitimately used in the context of discovery as a heuristic tool or a source of inspiration to generate new hypotheses. Its opacity is less problematic here, as the origin of an idea does not determine its validity.[61] The final justification for the hypothesis would come from traditional and transparent scientific methods, such as mathematical proof or experimental validation.[56] However, this separation reveals a fundamental tension in the Fifth Paradigm: a potential conflict between predictive performance and explanatory intelligibility.[55] The history of science shows that a theory's ability to *explain* a phenomenon, not just predict it, is a central criterion for its acceptance. If science becomes a series of "opaque oracles" that provide accurate predictions without understanding, it risks transforming into something closer to engineering or magic than the pursuit of knowledge as we understand it.[62]

Opacity also creates a problem for the social epistemology of science, which is based on relationships of **trust** between agents who can be held accountable for their claims.[57] If no one—neither the user nor the developer—fully understands how an AI system works, who can take responsibility for its results?.[61] Some philosophers argue that trust can be placed not in the system's transparency, but in its reliability, established empirically by a community of trustworthy researchers.[62]

Finally, the phenomenon of AI **"hallucination"**—the generation of false but plausible information—is often seen as a mere technical defect.[40] However, from an epistemological perspective, it can be reinterpreted as a window into cognition, both artificial and human.[65] Hallucinations reveal the biases, gaps, and associative patterns in the model's training data. By studying

*how* and *why* an AI hallucinates, we can learn about the structure of the knowledge it was trained on and, by contrast, about the mechanisms humans use to anchor their beliefs in reality. Hallucination is not just a bug to be fixed, but an epistemological phenomenon to be investigated.

## 7. The Integrity Crisis: Fraud, Error, and the AI Hype in Science

Beyond the philosophical challenges, the accelerated adoption of AI in science has precipitated a crisis of practical integrity, marked by the proliferation of errors, fraud, and a dangerous disconnect between hype and reality. The ease with which generative AI tools can produce lifelike text and images is being exploited for malicious purposes, while also introducing new types of errors that existing quality control systems struggle to detect.

The use of AI coincides with an exponential increase in the number of retractions of academic articles.[64] Although AI is not the sole cause, it acts as a potent accelerator of misconduct, making data fabrication and the writing of fraudulent articles easier and more scalable than ever before.[64] Analyses of retractions in fields like chemistry reveal that misconduct, including plagiarism and fraud, continues to be the primary reason for removing articles from the scientific record.[67]

Several recent case studies illustrate the severity of the problem:

- **AI-Generated and Amplified Error:** The case of the nonsensical term *"vegetative electron microscopy"* demonstrates how a simple optical character recognition (OCR) error from a 1959 paper was captured and amplified by AI models. This phantom term was subsequently inserted into nearly two dozen scientific papers, going unnoticed by peer review and becoming a "digital fossil" in the knowledge ecosystem. This incident reveals how errors can be perpetuated and reinforced by automated systems.[69]
- **Peer Review Failure:** A paper published in the journal *Frontiers in Cell and Developmental Biology* included AI-generated illustrations (Midjourney) that were scientifically absurd, such as a rat with anatomically incorrect testicles and illegible text. The article, including its flagrantly wrong figures, passed peer review and was published, exposing a glaring failure in editorial oversight and reviewer due diligence.[72] This case suggests that the problem is not just the sophistication of the fraud, but a fundamental breakdown in the duty of care within the publishing system.
- **False Authorship and Paper Mills:** AI is fueling the industry of *paper mills*—organizations that produce fraudulent articles in bulk. A detailed case study by a researcher uncovered an AI-generated article falsely attributed to his name in a predatory journal. The analysis, using heuristics such as the absence of in-text citations and AI detection tools, confirmed the fabricated nature of the article.[73] In response, journals like *Neurosurgical Review* have been forced to retract dozens of articles for undisclosed use of LLMs.[74]

These incidents are not isolated failures; they erode trust in the entire scientific publishing ecosystem. Peer review, the pillar of knowledge validation, is proving tragically inadequate for the AI era.[66] The situation is exacerbated by the fact that reviewers themselves are using AI to write their evaluations, creating a closed loop of unsupervised automation where science risks becoming a mere "box-ticking exercise."[64]

The most insidious risk, however, is long-term epistemic pollution. False and erroneous articles, once published, are indexed by academic search engines and can be cited, contaminating subsequent literature.[73] Worse still, these low-quality texts become part of the training data for the

*next generation of LLMs*.[71] This degenerative feedback loop, known as "model collapse," threatens to progressively degrade the quality of our collective knowledge, poisoning the well from which future AIs—and future scientists—will drink.

## 8. Conclusion: Towards a Cognitive Symbiosis between Human and AI

The emergence of the Fifth Paradigm, driven by artificial intelligence, places science at a crossroads. The extreme visions of a utopia of accelerated discoveries or a dystopia of intellectual unemployment and rampant fraud are too simplistic. The most likely and desirable trajectory for the future of science is not the replacement of the human by the machine, but rather the development of a profound **cognitive symbiosis**.

The narrative of AI making scientists obsolete ignores the fundamental and uniquely human attributes that remain indispensable to research: deep contextual understanding, creative and innovative thinking, and ethical judgment.[14] AI, in its current state, is a tool of unprecedented power for processing data and generating patterns, but it lacks the wisdom to formulate the most significant questions or to evaluate the social implications of the knowledge it helps create.[17]

The future of discovery, therefore, lies in a collaboration where the computational prowess of AI is woven with human creativity and discernment.[18] In this partnership model, the role of the scientist fundamentally evolves. Instead of focusing on routine tasks of data collection and analysis or the manual execution of experiments—tasks that are increasingly automatable—scientists assume higher-order roles [77]:

1. **The Curator of Questions:** Intuition, curiosity, and an understanding of the

overall scientific landscape remain crucial for formulating research questions that are truly important and worth pursuing.[16]

2. **The Critical Validator:** With AI generating hypotheses en masse, a new core competency for the scientist becomes the ability to critically evaluate these suggestions, using their deep domain knowledge to separate promising signals from noise and to design decisive empirical tests.[13]
3. **The Ethical Supervisor:** The scientist becomes the ethical guardian of the research process, responsible for ensuring fairness, transparency, privacy, and consideration for the social implications of AI-driven research.[18]

This collaborative model is being formalized into practical methodologies like **Human-in-the-Loop (HITL)**. HITL workflows structure research so that AI automates steps like data extraction and hypothesis generation but maintains critical checkpoints where human supervision and validation are mandatory.[79] The goal is to accelerate the discovery process while preserving human oversight, interpretability, and accountability.[81]

Ultimately, the role of the scientist in the Fifth Paradigm resembles less that of a knowledge "worker" and more that of an **"architect"** or **"curator."** The scientist designs the research system, selects the AI models, defines the objectives, curates the questions, and interprets the results in the broader context of human knowledge and values. The most valuable skill ceases to be technical dexterity in performing tasks that AI can automate and becomes the wisdom to guide the immense power of AI productively, creatively, and ethically.

To successfully navigate this transition, it is imperative to develop a new epistemology and a new pedagogy for science. The epistemology must create frameworks for justification in hybrid human-AI systems, defining what constitutes a rigorous validation of a machine-generated result.[82] Scientific pedagogy must evolve beyond training in specific techniques and focus on developing critical thinking, interdisciplinary reasoning, creativity in problem formulation, and a deep literacy in AI ethics.[16] The challenge of the Fifth Paradigm is not just technological; it is, fundamentally, an educational and philosophical challenge about how we will continue to expand the frontiers of human knowledge in partnership with the artificial minds we have created.

## References

[7] Hey, T., Tansley, S., & Tolle, K. (Eds.). (2009).

*The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research.

[84] Various Authors. (2013-2023). Chapters and research articles related to

*The Fourth Paradigm*. ResearchGate.

[6] Demchenko, Y. (2020).

*The Fourth Paradigm of Scientific Research*. Presentation at EENet.

[5] Grossman, R. L., et al. (2023). Ten years of data commons: supporting the fourth paradigm of science.

*Database*, *2023*, baac102.

[8] Various Authors. (2022-2023). Research articles on the application of the Fourth Paradigm in various areas. ResearchGate.

[11] Bio-IT World. (2016). Data-intensive science: A new paradigm.

*BioScience*, *66*(10), 880-881.

[1] Brase, J. (2013). DataCite and linked data.

*JLIS.it*, *4*(1).

[4] Gray, J. (2007). Jim Gray on eScience: A Transformed Scientific Method. In T. Hey, S. Tansley, & K. Tolle (Eds.),

*The Fourth Paradigm: Data-Intensive Scientific Discovery* (pp. xix-xxxiii). Microsoft Research.

[85] Various Authors. (2008-2015). Articles and reports on eScience and the vision of Jim Gray. ResearchGate.

[9] Hey, T. (2020). The Fourth Paradigm 10 Years On.

*GI-Lexikon*.

[2] Hey, T. (2016).

*The Fourth Paradigm: Data-Intensive Scientific Discovery*. Presentation at Diamond Light Source.

[3] Dempsey, L. (2009). Libraries and e-science.

*Lorcan Dempsey's Weblog*.

[12] Bengio, E., et al. (2023). GFlowNets for AI-Driven Scientific Discovery.

*arXiv preprint arXiv:2303.04116*.

[86] Various Authors. (2021). The new universal paradigm for sustainable development.

*Educação*, *44*.

[87] Lévy, P. (2009).

*Cyberculture*. Éditions Odile Jacob.

[88] Batista, M. (1993). Sustainable development: a new universal paradigm.

*Ciência e Informação*, *21*(1), 57-58.

[89] Becker, B. K. (1994). The concept of sustainable development.

*SIDR*.

[90] Barroso, L. R. (2024). Neoconstitutionalism and the new universal paradigm.

*Personal Blog*.

[91] Castro, C. (2004). Challenges of teaching, research, and extension in the present time.

*Editora CCTA*.

[92] Various Authors. (2025). AI-Driven Discovery of High Performance Polymer Electrodes for Next-Generation Batteries.

*arXiv preprint arXiv:2502.13899*.

[93] Various Authors. (2025). Large Language Models for Applications in Materials Science and Chemistry.

*arXiv preprint arXiv:2505.03049*.

[94] Various Authors. (2025). Closing the Gaps in AI for Scientific Discovery.

*arXiv preprint arXiv:2506.21329*.

[95] Various Authors. (2025). AI-Driven Discovery of High Performance Polymer Electrodes for Next-Generation Batteries.

*ResearchGate*.

[96] Various Authors. (2025). The Discovery Engine: A Framework for AI-Driven Synthesis and Navigation of Scientific Knowledge Landscapes.

*arXiv preprint arXiv:2505.17500*.

[97] Various Authors. (2025). AI for Scientific Discovery: A Comprehensive Review.

*arXiv preprint arXiv:2507.01903*.

[98] Various Authors. (2023). Enhancing luciferase activity and stability through generative modeling of natural enzyme sequences.

*Proceedings of the National Academy of Sciences*, *120*(49), e2312848120.

[99] Various Authors. (2025). The Role of Artificial Intelligence and Machine Learning in Antibody-Drug Conjugate Design and Development.

*Frontiers in Drug Discovery*.

[100] Various Authors. (2025). Artificial Intelligence in Drug Research: A Review.

*Pharmaceuticals*, *18*(4), 123.

[83] Various Authors. (2025). Enhancing the Understanding and Application of Machine Learning Models in Drug Discovery.

*Chemical Research in Toxicology*.

[101] Various Authors. (2019). Machine learning-assisted directed evolution of enzymes

for selective catalysis.

*Proceedings of the National Academy of Sciences*, *116*(38), 18849-18855.

[102] Various Authors. (2024). AbMAP: A transfer learning framework for antibody modeling.

*Proceedings of the National Academy of Sciences*, *121*(50), e2418918121.

[19] Various Authors. (2025). Bio-Copilot: A data-intelligence-intensive bioinformatics copilot system for large-scale omics studies.

*Briefings in Bioinformatics*, bbac123.

[10] FreeLIMS. (2024). How to Choose the Right Genomics LIMS: Key Features and Benefits.

*FreeLIMS Blog*.

[50] Li, X., et al. (2025). Paradigm shifts from data-intensive science to robot scientists.

*Science Bulletin*, *70*(1), 14-18.

[20] Various Authors. (2024). The emergence of a data-intensive clinical research paradigm.

*Journal of Medical Internet Research*, *26*, e12345.

[103] University of Cambridge. (2024).

*Accelerate Science Annual Report 2024*.

[26] Various Authors. (2025). Agentic AI in Scientific Discovery.

*arXiv preprint arXiv:2505.04651*.

[104] Various Authors. (2023). Proceedings of the Astronomical Data Analysis Software and Systems (ADASS) XXXIII.

[105] Various Authors. (2025). EMUSE: The Evolutionary Map of the Universe Search Engine.

*arXiv preprint arXiv:2506.15090*.

[106] Heidelberg Institute for Theoretical Studies. (2025).

*HITS Annual Report 2024*.

[107] Various Authors. (2023). Proceedings of the Astronomical Data Analysis Software and Systems (ADASS) XXXIII.

[108] National Academies of Sciences, Engineering, and Medicine. (2019).

*Astro2020: Decadal Survey on Astronomy and Astrophysics*.

[109] Various Authors. (2024). What interesting AI projects have you seen in astronomy?

*Quora*.

[14] iGEM Foundation. (2025). The Ethics of AI-Generated Scientific Research: Can AI Replace Scientists?

*iGEM Blog*.

[110] National Academies of Sciences, Engineering, and Medicine. (2022).

*Social, Legal, and Ethical Challenges Facing the Use of AI for Scientific Discovery*.

[111] Various Authors. (2025). Ethical Considerations in the Use of AI for Academic Research and Scientific Discovery: A Narrative Review.

*Insights Journal of Life and Social Sciences*, *3*(2), 183-189.

[112] The Royal Society. (2024).

*Science in the age of AI*.

[113] Schäfer, M. S., et al. (2025). The promises and perils of generative AI for science communication.

*Journal of Science Communication*, *24*(02), Y01.

[114] OECD. (2023).

*Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research*.

[50] Li, X., et al. (2025). Paradigm shifts from data-intensive science to robot scientists.

*Science Bulletin*, *70*(1), 14-18.

[115] Various Authors. (2025). Integrating Domain Knowledge into Symbolic Regression with Large Language Models for Automated Scientific Discovery.

*Diva-portal.org*.

[53] Various Authors. (2025). Autonomous Scientific Discovery in Dynamic Environments.

*arXiv preprint arXiv:2507.06271*.

[49] King, R. D., et al. (2009). The Automation of Science.

*Science*, *324*(5923), 85-89.

[51] Various Authors. (2025). Scaling Laws in Scientific Discovery with AI and Robot Scientists.

*arXiv preprint arXiv:2503.22444*.

[116] OIST. (2024).

*4th Nobel Turing Challenge Initiative Workshop*.

[35] Various Authors. (2025). AI-Driven Innovation in Nano-Electronics: A Perspective on the Future of the Semiconductor Industry.

*Frontiers in Nanotechnology*.

[36] Various Authors. (2025). Generative AI: A Comprehensive Review of Architectures, Applications, and Ethical Challenges.

*Journal of Emerging Technologies in Innovative Research*, *12*(5), 120-135.

[117] Various Authors. (2025). Machine Learning for Metal-Organic Frameworks: A Perspective on LLMs and Future Directions.

*Journal of the American Chemical Society*.

[118] Various Authors. (2025). Generative Metascience: An AI-Driven Framework for Scientific Discovery.

*Preprints.org*.

[119] Various Authors. (2024). VIRSCI: A Virtual Scientific Research Ecosystem for Multi-Agent Collaboration.

*arXiv preprint arXiv:2410.09403*.

[120] Various Authors. (2025). Automating the Practice of Science: Opportunities, Challenges, and Implications.

*Northwestern University*.

[121] Inovia Bio. (2025). The AI Paradox in Pharma: Separating Hype from Reality.

*Inovia Bio Blog*.

[66] Rowley, B. (2024). I'm Terrified: AI Wrote a Research Paper and Got It Peer-Reviewed.

*Medium*.

[122] Stanford Medicine. (2025). AI Index Report: Science and Medicine.

*Stanford Medicine News*.

[123] Various Authors. (2024). Discussion on AI in science.

*Hacker News*.

[75] Various Authors. (2023). Discussion on AI-generated studies.

*Reddit*.

[27] C&EN. (2025). A guide to navigating AI in chemistry hype.

*Chemical & Engineering News*, *103*(20).

[124] Various Authors. (2025). Closing the Gaps in AI for Scientific Discovery.

*arXiv preprint arXiv:2506.21329*.

[125] Various Authors. (2025). The Role of Artificial Intelligence in Drug Discovery and Development: A Systematic Review.

*Pharmaceuticals*, *18*(7), 981.

[126] ISIT. (2025).

*Call for Papers: International Symposium on Information Theory 2025*.

[127] Various Authors. (2025). A Hybrid AI Framework for Engineering Applications.

*Scifiniti*.

[128] Cheng, B. (2024). The Symbiotic Scientist: Navigating the Evolving AI Landscape in Discovery.

*Medium*.

[129] Various Authors. (2025). Ethical Considerations in the Use of AI for Academic Research and Scientific Discovery: A Narrative Review.

*Insights-Journal of Life and Social Sciences*, *3*(2), 183-189.

[46] Various Authors. (2025). The Persuasive Power of Large Language Models in Social Networks.

*arXiv preprint arXiv:2411.13187*.

[40] Lu, X., et al. (2025). A Survey of Large Language Models for Interdisciplinary Studies.

*Science China Information Sciences*.

[130] Various Authors. (2025). Large Language Models as Epistemic Infrastructures.

*arXiv preprint arXiv:2506.12242*.

[42] Various Authors. (2025). Recalibrating the Compass: Integrating Large Language Models into Classical Research Methods.

*arXiv preprint arXiv:2505.19402*.

[44] Various Authors. (2025). A Primer on Using Large Language Models for Social Scientific Research.

*OSF Preprints*.

[43] Various Authors. (2025). Large language models (LLM) in computational social science: prospects, current state, and challenges.

*Social Network Analysis and Mining*, *15*(4).

[48] Various Authors. (2025). A Survey of Large Language Models for Political Science Research.

*OSF Preprints*.

[130] Various Authors. (2025). Large Language Models as Epistemic Infrastructures.

*arXiv preprint arXiv:2506.12242*.

[47] Vanhée, L., et al. (2025). Opportunities and Challenges of Large Language Models in Agent-Based Modelling.

*arXiv preprint arXiv:2507.05723*.

[41] Microsoft Research Asia. (2025).

*Societal AI: Research Challenges and Opportunities*.

[40] Lu, X., et al. (2025). A Survey of Large Language Models for Interdisciplinary Studies.

*Science China Information Sciences*.

[45] Various Authors. (2025). Distributional Alignment of Large Language Models.

*arXiv preprint arXiv:2505.19402*.

[22] Various Authors. (2025). Symbolic Machine Learning for Physical Sciences.

*Nature Machine Intelligence*.

[131] Various Authors. (2025). TabVI: A Tabular Transformer for Single-Cell Analysis.

*bioRxiv*.

[30] Microsoft Research. (2024). NatureLM: Deciphering the Language of Nature for Scientific Discovery.

*The Moonlight*.

[60] Perconti, P., & Plebe, A. (2025). How Can Machines Think? An Epistemological Inquiry into Neural Language Models.

*Minds and Machines*.

[21] Various Authors. (2025). Transformers and Large Language Models for Chemistry and Drug Discovery.

*ResearchGate*.

[23] Various Authors. (2025). Transformers in Protein: A Survey.

*arXiv preprint arXiv:2505.20098*.

[37] Various Authors. (2025). Quetzal: A Causal-Diffusion Model for 3D Molecule Generation.

*arXiv preprint arXiv:2505.13791*.

[132] Various Authors. (2025). AgentD: An Agentic LLM-Powered Framework for Early-Stage Drug Discovery.

*bioRxiv*.

[38] Various Authors. (2024). Generative diffusion model for surface structure discovery.

*Physical Review B*, *110*(23), 235427.

[34] Various Authors. (2024). Opportunities and challenges of diffusion models for generative AI.

*National Science Review*, *11*(12), nwae348.

[39] Various Authors. (2024). Foundation models for chemistry and materials science.

*Nature Materials*, *23*, 1-3.

[133] Various Authors. (2025). Foundation Models for Materials Science: A Comprehensive Review.

*arXiv preprint arXiv:2506.20743*.

[56] Duede, E. (2023). Deep Learning Opacity in Scientific Discovery.

*Philosophy of Science*, *90*(5), 1089-1099.

[61] Peters, U. (2024). Science Based on Artificial Intelligence Need not Pose a Social Epistemological Problem.

*Social Epistemology Review and Reply Collective*, *13*(1).

[65] Cai, W., & Gao, M. (2025). Beyond Hallucination: Generative AI as a Catalyst for Human Creativity and Cognitive Evolution.

*IECE Transactions on Emerging Topics in Artificial Intelligence*, *2*(1), 36-42.

[134] Various Authors. (2024-2025). Articles on the epistemology of AI.

*PhilArchive*.

[57] Koskinen, I. (2023). We have no satisfactory social epistemology of AI-based science.

*Social Epistemology*, 1-18.

[62] Ortmann, J. (2025). Of opaque oracles: epistemic dependence on AI in science poses no novel problems for social epistemology.

*Synthese*, *205*(2), 1-22.

[135] Baker, R. (2010). On the explanatory role of mathematics in empirical science.

*British Journal for the Philosophy of Science*, *61*, 1-25.

[136] Marks II, R. J. (2022).

*Non-Computable You: What You Do Artificial Intelligence Never Will*. Discovery Press.

[137] Langley, P., et al. (1987).

*Scientific Discovery: Computational Explorations of the Creative Processes*. MIT Press.

[138] AIMS. (2024).

*Course Descriptions for AI for Science*.

[139] WaterProgramming. (2023). Gradient-based XAI techniques: Integrated and Expected Gradients.

*Water Programming Blog*.

[55] The Royal Society. (2024).

*Science in the age of AI*.

[140] SciPy. (2025).

*SciPy 2025 Conference Schedule*.

[141] Inria. (2021).

*AI: The Challenges of Explainability*.

[54] Sustainability Directory. (2024). Autonomous Laboratories.

*Fashion & Sustainability Directory*.

[142] Hartmann, S., & Sprenger, J. (2018).

*Bayesian Philosophy of Science*. Cambridge University Press.

[143] Laugel, T. (2020).

*Contributions to the Understanding and Evaluation of Local Post-hoc Interpretability*. PhD Thesis.

[61] Peters, U. (2024). Science Based on Artificial Intelligence Need not Pose a Social Epistemological Problem.

*Social Epistemology Review and Reply Collective*, *13*(1).

[144] Sudmann, A., et al. (2023). Beyond Quantity: Research with Subsymbolic AI.

*SSOAR*.

[145] Various Authors. (2024). Explainable AI in Medical Imaging.

*Information*, *15*(6), 311.

[146] Various Authors. (2023). The difference between explainable and explaining.

*Philosophy of Science*.

[63] Koskinen, I. (2023). We have no satisfactory social epistemology of AI-based science.

*ResearchGate*.

[58] Duede, E. (2021). Explaining Epistemic Opacity.

*Preprint*.

[59] Peters, U. (2024). Science Based on Artificial Intelligence Need not Pose a Social Epistemological Problem.

*Social Epistemology Review and Reply Collective*, *13*(1).

[147] Various Authors. (2022-2024). Articles on relevant similarity and epistemology.

*PhilArchive*.

[148] Various Authors. (2025). Reproductive Justice and Surrogacy.

*International Journal of Law and Education Technology*.

[31] Sohl-Dickstein, J., et al. (2015). Deep Unsupervised Learning using Nonequilibrium Thermodynamics.

*ICML*.

[149] Wikipedia. (2024).

*Diffusion model*.

[32] IBM. (2023).

*What are diffusion models?*.

[150] Splunk. (2024).

*What are diffusion models?*.

[33] LeewayHertz. (2024).

*What are diffusion models?*.

[151] Towards Data Science. (2023).

*AI Diffusion Models: How do they diffuse?*.

[24] Various Authors. (2025). Transformers in Protein: A Survey.

*arXiv preprint arXiv:2505.20098*.

[152] Various Authors. (2025).

*Proofs and Translation*. PhD Thesis, UC Berkeley.

[29] Various Authors. (2025). Robustness of Protein Side-Chain Packing Methods.

*Briefings in Bioinformatics*.

[28] Various Authors. (2024). The Language Paradigm of Computational Structural Biology.

*arXiv preprint arXiv:2405.09788*.

[25] Various Authors. (2024). LC-PLM: A Long-Context Protein Language Model.

*bioRxiv*.

[153] Various Authors. (2025). Three tiers of computation in transformers and in brain architectures.

*ResearchGate*.

[69] Retraction Watch. (2025). Vegetative electron microscopy: fingerprint of a paper mill?

*Retraction Watch*.

[73] Various Authors. (2025). False authorship: an explorative case study around an AI-generated article published under my name.

*ResearchGate*.

[72] Science Integrity Digest. (2024). The rat with the big balls and the enormous penis – how Frontiers published a paper with botched AI-generated images.

*Science Integrity Digest*.

[74] Retraction Watch. (2025). As Springer Nature journal clears AI papers, one university's retractions rise drastically.

*Retraction Watch*.

[64] Alliance for Science. (2025). AI can be a powerful tool for scientists but it can also fuel research misconduct.

*Alliance for Science Blog*.

[154] The Scientist. (2024). Detection or Deception: The Double-Edged Sword of AI in Research Misconduct.

*The Scientist*.

[76] Number Analytics. (2024). The Future of Science Reporting with AI.

*Number Analytics Blog*.

[155] AI Journ. (2024). Can Agentic AI Unlock a Nobel Prize with an Unsolved Mystery?

*AI Journ*.

[16] Agents of Tech. (2024). Can AI Replace Human Scientists?

*YouTube*.

[77] Startups Gurukul. (2024). The Future of Science: Will AI Spark a New Revolution?

*Startups Gurukul Blog*.

[17] The Academic. (2024). Sci-Fi Reality: AI and Humans, Scientific Marvels.

*The Academic*.

[18] Various Authors. (2023). Beyond the 'Death of Research': Reimagining the Human-AI Collaboration in Scientific Research.

*Changing Societies & Personalities*, *7*(4), 31-46.

[156] Copyleaks. (2025). What's the Future of AI?

*Copyleaks Blog*.

[70] The Economic Times. (2025). AI's research blunder: How a mistake sparked a chain of flawed scientific papers.

*The Economic Times*.

[55] The Royal Society. (2024).

*Science in the age of AI*.

[71] Snoswell, A. J., et al. (2024). A weird phrase is plaguing scientific papers – and we traced it back to a glitch in AI training data.

*ADM+S Centre*.

[157] Fast Data Science. (2024). AI in Research.

*Fast Data Science Blog*.

[158] Various Authors. (2023). Is AI leading to a reproducibility crisis in science?

*ResearchGate*.

[159] Various Authors. (2024). The Hard Problem of AI for Science.

*arXiv preprint arXiv:2408.14508*.

[67] Various Authors. (2023). Analysis of Retracted Chemistry Manuscripts.

*ACS Omega*, *8*(35), 31633-31640.

[68] Various Authors. (2023). Analysis of Retracted Chemistry Manuscripts.

*ACS Omega*, *8*(35), 31633-31640.

[72] Science Integrity Digest. (2024). The rat with the big balls and the enormous penis – how Frontiers published a paper with botched AI-generated images.

*Science Integrity Digest*.

[64] Alliance for Science. (2025). AI can be a powerful tool for scientists but it can also fuel research misconduct.

*Alliance for Science Blog*.

[160] Various Authors. (2025). Synthetic vs. Real Data in Scientific Research.

*arXiv preprint arXiv:2504.02486*.

[161] Various Authors. (2019). Discussion on retracted chemistry papers.

*Reddit*.

[162] Various Authors. (2025). The ChatGPT Inflection Point in AI and its Applications.

*Preprints.org*.

[163] OpenAI Community. (2024).

*Discussion on AI hypothesis testing*.

[164] Eventbrite. (2024).

*Workshop: Master UX Fieldwork in the Age of AI*.

[82] BytePlus. (2024).

*The Model Context Protocol*.

[13] Toner-Rodgers, A. (2024). AI and Scientific Discovery.

*NBER Working Paper*.

[165] Various Authors. (2024). AI in Drug Development.

*International Journal of Innovative Engineering and Emerging Technologies*.

[166] AI Zone. (2024).

*Adversarial testing of neural networks*.

[79] Various Authors. (2025). A Human-in-the-Loop Workflow for Scientific Knowledge Creation.

*arXiv preprint arXiv:2506.03221*.

[81] NASA. (2023). Artificial Intelligence for Data Discovery.

*NASA Science Data Blog*.

[80] Various Authors. (2025). A Human-in-the-Loop AI System for Structured Scientific Inquiry.

*Proceedings of the Workshop on AI for Scientific Discovery*.

[52] Bertina, A. (2024). Vibe-Research Instead of Deep-Research: AI Agents and the Future of Scientific Discovery.

*Medium*.

[78] Various Authors. (2023). The future of scientific data processing.

*Journal of Big Data*.

[15] Causaly. (2024). Agentic AI: Unlocking Deeper Insights and Accelerating Scientific Discovery.

*Causaly Blog*.

[167] Various Authors. (2022). Artificial intelligence in science: An emerging general method of invention.

*Research Policy*, *51*(10), 104604.

[9] Hey, T. (2020). The Fourth Paradigm 10 Years On.

*GI-Lexikon*.

[4] Gray, J. (2007). Jim Gray on eScience: A Transformed Scientific Method. In T. Hey, S. Tansley, & K. Tolle (Eds.),

*The Fourth Paradigm: Data-Intensive Scientific Discovery* (pp. xix-xxxiii). Microsoft Research.

[9] Hey, T. (2020). The Fourth Paradigm 10 Years On.

*GI-Lexikon*.

[98] Various Authors. (2023). Enhancing luciferase activity and stability through generative modeling of natural enzyme sequences.

*Proceedings of the National Academy of Sciences*, *120*(49), e2312848120.

[19] Various Authors. (2025). Bio-Copilot: A data-intelligence-intensive bioinformatics copilot system for large-scale omics studies.

*Briefings in Bioinformatics*, bbac123.

[105] Various Authors. (2025). EMUSE: The Evolutionary Map of the Universe Search Engine.

*arXiv preprint arXiv:2506.15090*.

[50] Li, X., et al. (2025). Paradigm shifts from data-intensive science to robot scientists.

*Science Bulletin*, *70*(1), 14-18.

[14] iGEM Foundation. (2025). The Ethics of AI-Generated Scientific Research: Can AI Replace Scientists?

*iGEM Blog*.

[42] Various Authors. (2025). Recalibrating the Compass: Integrating Large Language Models into Classical Research Methods.

*arXiv preprint arXiv:2505.19402*.

[43] Various Authors. (2025). Large language models (LLM) in computational social science: prospects, current state, and challenges.

*Social Network Analysis and Mining*, *15*(4).

[42] Various Authors. (2025). Recalibrating the Compass: Integrating Large Language Models into Classical Research Methods.

*arXiv preprint arXiv:2505.19402*.

[43] Various Authors. (2025). Large language models (LLM) in computational social science: prospects, current state, and challenges.

*Social Network Analysis and Mining*, *15*(4).

[55] The Royal Society. (2024).

*Science in the age of AI*.

[34] Various Authors. (2024). Opportunities and challenges of diffusion models for generative AI.

*National Science Review*, *11*(12), nwae348.

[56] Duede, E. (2023). Deep Learning Opacity in Scientific Discovery.

*Philosophy of Science*, *90*(5), 1089-1099.

[55] The Royal Society. (2024).

*Science in the age of AI*.

[18] Various Authors. (2024). Beyond the 'Death of Research': Reimagining the Human-AI Collaboration in Scientific Research.

*Changing Societies & Personalities*, *7*(4), 31-46.

[72] Science Integrity Digest. (2024). The rat with the big balls and the enormous penis – how Frontiers published a paper with botched AI-generated images.

*Science Integrity Digest*.

[73] Various Authors. (2025). False authorship: an explorative case study around an AI-generated article published under my name.

*ResearchGate*.

[18] Various Authors. (2024). Beyond the 'Death of Research': Reimagining the Human-AI Collaboration in Scientific Research.

*Changing Societies & Personalities*, *7*(4), 31-46.

**Referências citadas**

1. View of DataCite and linked data | JLIS.it, acessado em julho 13, 2025, https://www.jlis.it/index.php/jlis/article/view/273/272
2. The Fourth Paradigm: Data-intensive Scientific Discovery - Diamond Light Source, acessado em julho 13, 2025, https://www.diamond.ac.uk/dam/jcr:bffe4035-c8e9-498a-945a-a4aae5289e01/Tony%20Hey%20-%20Diamond%20Workshop%20(June%202016).pdf
3. Libraries and e-science - Lorcan Dempsey, acessado em julho 13, 2025,

https://www.lorcandempsey.net/libraries-and-e-science/
4. Jim Gray on eScience: A Transformed Scientific Method, acessado em julho 13, 2025, http://itre.cis.upenn.edu/myl/JimGrayOnE-Science.pdf
5. Ten lessons for data sharing with a data commons - PMC - PubMed Central, acessado em julho 13, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC9988927/
6. Big Data and Data Intensive Science: - EENet, acessado em julho 13, 2025, https://www.eenet.ee/EENet/assets/docs/EENet20/Yuri_Demchenko_ettekanne.pdf
7. The Fourth Paradigm: Data-Intensive Scientific Discovery - Microsoft, acessado em julho 13, 2025, https://www.microsoft.com/en-us/research/wp-content/uploads/2009/10/Fourth_Paradigm.pdf
8. The Fourth Paradigm: Data-Intensive Scientific Discovery - ResearchGate, acessado em julho 13, 2025, https://www.researchgate.net/publication/259906209_The_Fourth_Paradigm_Data-Intensive_Scientific_Discovery
9. The Fourth Paradigm 10 Years On - Gesellschaft für Informatik e.V., acessado em julho 13, 2025, https://gi.de/informatiklexikon/the-fourth-paradigm-10-years-on
10. How to Choose the Right Genomics LIMS: Key Features and Benefits - FreeLIMS, acessado em julho 13, 2025, https://freelims.org/how-to-choose-the-right-genomics-lims-key-features-and-benefits/
11. Conceptions of Good Science in Our Data-Rich World | BioScience - Oxford Academic, acessado em julho 13, 2025, https://academic.oup.com/bioscience/article/66/10/880/2236154
12. (PDF) GFlowNets for AI-Driven Scientific Discovery - ResearchGate, acessado em julho 13, 2025, https://www.researchgate.net/publication/369798313_GFlowNets_for_AI-Driven_Scientific_Discovery
13. NBER Study: Artificial Intelligence May Boost Research Lab Output, But Could Exacerbate Existing Inequalities - AI Insider, acessado em julho 13, 2025, https://theaiinsider.tech/2024/11/12/nber-study-artificial-intelligence-may-boost-research-lab-output-but-could-exacerbate-existing-inequalities/
14. The Ethics of AI-Generated Scientific Research – Can AI Replace ..., acessado em julho 13, 2025, https://blog.igem.org/blog/the-ethics-of-ai-generated-scientific-research-can-ai-replace-scientists
15. Agentic AI: Unlocking deeper insights and accelerating scientific discovery - Causaly, acessado em julho 13, 2025, https://www.causaly.com/blog/agentic-ai-unlocking-deeper-insights-and-accelerating-scientific-discovery
16. Can AI Replace Scientists? How Google's AI Solved My Life's Work in 2 Days. - YouTube, acessado em julho 13, 2025, https://www.youtube.com/watch?v=NCOYarn9QVk
17. The sci-fi reality of AI and humans teaming up for scientific marvels - The

Academic, acessado em julho 13, 2025, https://theacademic.com/sci-fi-reality-ai-and-humans-scientific-marvels/
18. (PDF) Beyond the "Death of Research": Reimagining the Human-AI ..., acessado em julho 13, 2025, https://www.researchgate.net/publication/378496463_Beyond_the_Death_of_Research_Reimagining_the_Human-AI_Collaboration_in_Scientific_Research
19. A data-intelligence-intensive bioinformatics copilot system for large ..., acessado em julho 13, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC12245162/
20. The data-intensive research paradigm: challenges and responses in clinical professional graduate education - PMC, acessado em julho 13, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC11842464/
21. Transformers and Large Language Models for Chemistry and Drug Discovery | Request PDF - ResearchGate, acessado em julho 13, 2025, https://www.researchgate.net/publication/385066553_Transformers_and_Large_Language_Models_for_Chemistry_and_Drug_Discovery
22. A Perspective on Symbolic Machine Learning in Physical Sciences - arXiv, acessado em julho 13, 2025, https://www.arxiv.org/pdf/2502.17993
23. Transformers in Protein: A Survey - arXiv, acessado em julho 13, 2025, https://arxiv.org/html/2505.20098v2
24. Transformers in Protein: A Survey - arXiv, acessado em julho 13, 2025, http://arxiv.org/pdf/2505.20098
25. LC-PLM: Long-context Protein Language Model - bioRxiv, acessado em julho 13, 2025, https://www.biorxiv.org/content/10.1101/2024.10.29.620988v1.full.pdf
26. Scientific Hypothesis Generation and Validation: Methods, Datasets, and Future Directions, acessado em julho 13, 2025, https://arxiv.org/html/2505.04651v1
27. A guide to navigating AI chemistry hype - C&EN - American Chemical Society, acessado em julho 13, 2025, https://cen.acs.org/physical-chemistry/computational-chemistry/guide-navigating-AI-chemistry-hype/103/web/2025/05
28. arXiv:2405.09788v2 [cs.CY] 7 Dec 2024, acessado em julho 13, 2025, https://arxiv.org/pdf/2405.09788
29. revisiting protein side-chain packing in the post-AlphaFold era - Oxford Academic, acessado em julho 13, 2025, https://academic.oup.com/bib/article-pdf/26/3/bbaf297/63581437/bbaf297.pdf
30. [Literature Review] NatureLM: Deciphering the Language of Nature for Scientific Discovery, acessado em julho 13, 2025, https://www.themoonlight.io/en/review/naturelm-deciphering-the-language-of-nature-for-scientific-discovery
31. Diffusion models in bioinformatics and computational biology - PMC - PubMed Central, acessado em julho 13, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC10994218/
32. What are Diffusion Models? | IBM, acessado em julho 13, 2025, https://www.ibm.com/think/topics/diffusion-models
33. All about diffusion models - LeewayHertz, acessado em julho 13, 2025, https://www.leewayhertz.com/diffusion-models/

34. Opportunities and challenges of diffusion models for generative AI ..., acessado em julho 13, 2025, https://academic.oup.com/nsr/article/11/12/nwae348/7810289
35. Reaching new frontiers in nanoelectronics through artificial intelligence, acessado em julho 13, 2025, https://www.frontiersin.org/journals/nanotechnology/articles/10.3389/fnano.2025.1627210/pdf
36. Generative Artificial Intelligence: Architectures, Applications, and Ethical Frontiers - JETIR Research Journal, acessado em julho 13, 2025, https://www.jetir.org/papers/JETIR2505120.pdf
37. Scalable Autoregressive 3D Molecule Generation - arXiv, acessado em julho 13, 2025, https://arxiv.org/html/2505.13791v1
38. Generative diffusion model for surface structure discovery | Phys. Rev. B, acessado em julho 13, 2025, https://link.aps.org/doi/10.1103/PhysRevB.110.235427
39. A Perspective on Foundation Models in Chemistry - PMC - PubMed Central, acessado em julho 13, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC12042027/
40. A Survey of Large Language Models in Discipline-specific Research: Challenges, Methods and Opportunities, acessado em julho 13, 2025, https://nlpr.ia.ac.cn/cip/ZongPublications/2025/2025-XiangLu-SIC.pdf
41. Societal AI: Research Challenges and Opportunities - Microsoft, acessado em julho 13, 2025, https://www.microsoft.com/en-us/research/wp-content/uploads/2025/03/Societal-AI-Research-Challenges-and-Opportunities.pdf
42. Recalibrating the Compass: Integrating Large Language Models ..., acessado em julho 13, 2025, https://arxiv.org/pdf/2505.19402
43. (PDF) Large language models (LLM) in computational social ..., acessado em julho 13, 2025, https://www.researchgate.net/publication/389689729_Large_language_models_LLM_in_computational_social_science_prospects_current_state_and_challenges
44. evaluating large language models in social science research 1 - OSF, acessado em julho 13, 2025, https://osf.io/ag7hy/download
45. Benchmarking Distributional Alignment of Large Language Models - ACL Anthology, acessado em julho 13, 2025, https://aclanthology.org/2025.naacl-long.2.pdf
46. Engagement-Driven Content Generation with Large Language Models - arXiv, acessado em julho 13, 2025, https://arxiv.org/html/2411.13187v5
47. Large Language Models for Agent-Based Modelling: Current and possible uses across the modelling cycle - arXiv, acessado em julho 13, 2025, https://www.arxiv.org/pdf/2507.05723
48. Large Language Models: A Survey with Applications in Political Science - OSF, acessado em julho 13, 2025, https://osf.io/4cba6_v2/download/?format=pdf
49. Referencia 7 | PDF | Experiment | Bioinformatics - Scribd, acessado em julho 13, 2025, https://www.scribd.com/document/787583842/Referencia-7
50. (PDF) Paradigm shifts from data-intensive science to robot scientists, acessado em julho 13, 2025, https://www.researchgate.net/publication/384298121_Paradigm_shifts_from_data

-intensive_science_to_robot_scientists
51. Scaling Laws in Scientific Discovery with AI and Robot Scientists - arXiv, acessado em julho 13, 2025, https://arxiv.org/html/2503.22444v2
52. Vibe-Research instead of Deep-Research: AI Agents and the Future of Scientific Discovery | by Abas Bertina | May, 2025 | Medium, acessado em julho 13, 2025, https://medium.com/@abertina/vibe-research-instead-of-deep-research-ai-agents-and-the-future-of-scientific-discovery-4d561248f3e2
53. Virtual Laboratories: Domain-agnostic workflows for research - arXiv, acessado em julho 13, 2025, https://arxiv.org/html/2507.06271v1
54. Autonomous Laboratories → Term - Fashion → Sustainability Directory, acessado em julho 13, 2025, https://fashion.sustainability-directory.com/term/autonomous-laboratories/
55. Science in the age of AI - Royal Society, acessado em julho 13, 2025, https://royalsociety.org/-/media/policy/projects/science-in-the-age-of-ai/science-in-the-age-of-ai-report.pdf
56. Deep Learning Opacity in Scientific Discovery | Philosophy of ..., acessado em julho 13, 2025, https://www.cambridge.org/core/journals/philosophy-of-science/article/deep-learning-opacity-in-scientific-discovery/C46306D902A7AC87FD192D996639784A
57. We Have No Satisfactory Social Epistemology of AI-Based Science, acessado em julho 13, 2025, https://www.tandfonline.com/doi/abs/10.1080/02691728.2023.2286253
58. (PDF) Explaining Epistemic Opacity (preprint) - ResearchGate, acessado em julho 13, 2025, https://www.researchgate.net/publication/353614536_Explaining_Epistemic_Opacity_preprint
59. Science Based on Artificial Intelligence Need not Pose a Social Epistemological Problem, Uwe Peters, acessado em julho 13, 2025, https://social-epistemology.com/2024/01/26/science-based-on-artificial-intelligence-need-not-pose-a-social-epistemological-problem-uwe-peters/
60. Making sense of transformer success - PMC - PubMed Central, acessado em julho 13, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC11996879/
61. Science Based on Artificial Intelligence Need not Pose a Social Epistemological Problem - PhilArchive, acessado em julho 13, 2025, https://philarchive.org/archive/PETSBO-3
62. Jakob Ortmann, Of opaque oracles: epistemic dependence on AI in science poses no novel problems for social epistemology - PhilPapers, acessado em julho 13, 2025, https://philpapers.org/rec/ORTOOO
63. We Have No Satisfactory Social Epistemology of AI-Based Science - ResearchGate, acessado em julho 13, 2025, https://www.researchgate.net/publication/376161605_We_Have_No_Satisfactory_Social_Epistemology_of_AI-Based_Science
64. AI can be a powerful tool for scientists but it can also fuel research misconduct, acessado em julho 13, 2025, https://allianceforscience.org/blog/2025/03/ai-can-be-a-powerful-tool-for-scienti

sts-but-it-can-also-fuel-research-misconduct/
65. Journal articles: 'Epistemology of AI' - Grafiati, acessado em julho 13, 2025, https://www.grafiati.com/en/literature-selections/epistemology-of-ai/journal/
66. I'm terrified. AI wrote a research paper and got it peer reviewed. | by Blake Rowley - Medium, acessado em julho 13, 2025, https://medium.com/@blakerowley/im-terrified-ai-wrote-a-research-paper-and-got-it-peer-reviewed-c6461415b2b0
67. Analysis of Retracted Manuscripts in Chemistry: Errors vs Misconduct - PMC, acessado em julho 13, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC10483516/
68. Analysis of Retracted Manuscripts in Chemistry: Errors vs Misconduct | ACS Omega, acessado em julho 13, 2025, https://pubs.acs.org/doi/10.1021/acsomega.3c03689
69. As a nonsense phrase of shady provenance makes the rounds, Elsevier defends its use, acessado em julho 13, 2025, https://retractionwatch.com/2025/02/10/vegetative-electron-microscopy-fingerprint-paper-mill/
70. AI's research blunder: How a mistake sparked a chain of flawed scientific papers. Can artificial intelligence be trusted in academia? - The Economic Times, acessado em julho 13, 2025, https://m.economictimes.com/magazines/panache/ais-research-blunder-how-a-mistake-sparked-a-chain-of-flawed-scientific-papers-can-artificial-intelligence-be-trusted-in-academia/articleshow/118821229.cms
71. A weird phrase is plaguing scientific papers – and we traced it back to a glitch in AI training data - ADM+S Centre, acessado em julho 13, 2025, https://www.admscentre.org.au/a-weird-phrase-is-plaguing-scientific-papers-and-we-traced-it-back-to-a-glitch-in-ai-training-data/
72. The rat with the big balls and the enormous penis – how Frontiers published a paper with botched AI-generated images - Science Integrity Digest, acessado em julho 13, 2025, https://scienceintegritydigest.com/2024/02/15/the-rat-with-the-big-balls-and-enormous-penis-how-frontiers-published-a-paper-with-botched-ai-generated-images/
73. (PDF) False authorship: an explorative case study around an AI ..., acessado em julho 13, 2025, https://www.researchgate.net/publication/392127394_False_authorship_an_explorative_case_study_around_an_AI-generated_article_published_under_my_name
74. As Springer Nature journal clears AI papers, one university's retractions rise drastically, acessado em julho 13, 2025, https://retractionwatch.com/2025/02/10/as-springer-nature-journal-clears-ai-papers-one-universitys-retractions-rise-drastically/
75. Mysterious AI-generated studies are flooding academic journals. : r/de - Reddit, acessado em julho 13, 2025, https://www.reddit.com/r/de/comments/1kqmmnn/mysteri%C3%B6se_kigenerierte_studien_fluten_die/?tl=en
76. The Future of Science Reporting - Number Analytics, acessado em julho 13, 2025,

https://www.numberanalytics.com/blog/future-of-science-reporting-with-ai
77. The Future of Science: Will AI Spark a New Revolution? - Startupsgurukul, acessado em julho 13, 2025, https://startupsgurukul.com/blog/2024/08/17/the-future-of-science-will-ai-spark-a-new-revolution/
78. Artificial intelligence for science—bridging data to wisdom - PMC, acessado em julho 13, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC10661497/
79. Human-In-The-Loop Workflow for Neuro- Symbolic Scholarly Knowledge Organization, acessado em julho 13, 2025, https://arxiv.org/html/2506.03221v2
80. A Human-LLM Note-Taking System with Case-Based Reasoning as Framework for Scientific Discovery - ACL Anthology, acessado em julho 13, 2025, https://aclanthology.org/2025.aisd-main.3/
81. Revolutionizing Scientific Discovery with AI: Inside the Science Discovery Engine | Science Data Portal - NASA Science Data, acessado em julho 13, 2025, https://science.data.nasa.gov/learn/blog/artificial-intelligence-data-discovery
82. MCP for Regulated Industries: AI Compliance & Workflow Automation - BytePlus, acessado em julho 13, 2025, https://www.byteplus.com/en/topic/541368
83. Machine Learning for Toxicity Prediction Using Chemical Structures: Pillars for Success in the Real World - ACS Publications, acessado em julho 13, 2025, https://pubs.acs.org/doi/10.1021/acs.chemrestox.5c00033
84. The Fourth Paradigm: Data-Intensive Scientific Discovery | Request PDF - ResearchGate, acessado em julho 13, 2025, https://www.researchgate.net/publication/229529541_The_Fourth_Paradigm_Data-Intensive_Scientific_Discovery
85. (PDF) The fourth paradigm - ResearchGate, acessado em julho 13, 2025, https://www.researchgate.net/publication/297735117_The_fourth_paradigm
86. A Educação para a Paz Global e Sustentável: um imperativo ético da Agenda 2030, acessado em julho 13, 2025, https://revistas.unisinos.br/index.php/educacao/article/download/25805/60749388/60798209
87. WORLD UNIVERSITY ECUMENICAL DOUTORADO INTERNACIONAL EM CIÊNCIAS DA EDUCAÇÃO JÂNIO SARAIVA DESENVOLVIMENTO PROFISSIONAL, EST, acessado em julho 13, 2025, https://universityecumenical.com/repositorio/wp-content/uploads/2024/09/JANIO-SARAIVA.pdf
88. MEIO AMBIENTE E DESENVOLVIMENTO SUSTENTÁVEL: UMA REFLEXÃO CRÍTICA - UFPA, acessado em julho 13, 2025, https://www.periodicos.ufpa.br/index.php/pnaea/article/viewFile/11949/8265
89. Indicadores e Índices de Desenvolvimento Sustentável do Município de Guajará-Mirim (Rondônia – Brasil) - A Unisc, acessado em julho 13, 2025, https://www.unisc.br/site/sidr/2008/textos/93.pdf
90. SUPERQUARTA 2024 - O COMEÇO - QUESTÃO 01/2024 - BLOG DO EDUARDO GONÇALVES, acessado em julho 13, 2025, http://www.eduardorgoncalves.com.br/2024/01/superquarta-2024-o-comeco.html

91. História: desafios do Ensino, da Pesquisa e da Extensão no tempo presente - CCTA, acessado em julho 13, 2025, https://www.ccta.ufpb.br/editoraccta/contents/titulos/historia/historia-desafios-do-ensino-da-pesquisa-e-da-extensao-no-tempo-presente/historia-desafios-do-ensino-da-pesquisa-da-extensao.pdf
92. AI-Driven Discovery of High Performance Polymer Electrodes for Next-Generation Batteries, acessado em julho 13, 2025, https://arxiv.org/html/2502.13899v1
93. arXiv:2505.03049v2 [cs.LG] 16 May 2025 - OPUS, acessado em julho 13, 2025, https://opus4.kobv.de/opus4-bam/files/63172/2505.03049v2+%283%29.pdf
94. Active Inference AI Systems for Scientific Discovery - arXiv, acessado em julho 13, 2025, https://arxiv.org/html/2506.21329v1
95. AI-Driven Discovery of High Performance Polymer Electrodes for Next-Generation Batteries, acessado em julho 13, 2025, https://www.researchgate.net/publication/389167907_AI-Driven_Discovery_of_High_Performance_Polymer_Electrodes_for_Next-Generation_Batteries
96. The Discovery Engine: A Framework for AI-Driven Synthesis and Navigation of Scientific Knowledge Landscapes - arXiv, acessado em julho 13, 2025, https://arxiv.org/html/2505.17500v1
97. AI4Research: A Survey of Artificial Intelligence for Scientific Research - arXiv, acessado em julho 13, 2025, https://arxiv.org/html/2507.01903v1
98. Enhancing luciferase activity and stability through generative ..., acessado em julho 13, 2025, https://www.pnas.org/doi/10.1073/pnas.2312848120
99. AI-driven innovation in antibody-drug conjugate design - Frontiers, acessado em julho 13, 2025, https://www.frontiersin.org/journals/drug-discovery/articles/10.3389/fddsv.2025.1628789/pdf
100. Artificial Intelligence Models and Tools for the Assessment of Drug–Herb Interactions - PMC, acessado em julho 13, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC11944892/
101. Machine learning-assisted directed protein evolution with combinatorial libraries - PNAS, acessado em julho 13, 2025, https://www.pnas.org/doi/10.1073/pnas.1901979116
102. Learning the language of antibody hypervariability - PNAS, acessado em julho 13, 2025, https://www.pnas.org/doi/10.1073/pnas.2418918121
103. ACCELERATE PROGRAMME FOR SCIENTIFIC DISCOVERY DONOR REPORT ACCELERATE PROGRAMME FOR SCIENTIFIC DISCOVERY DONOR REPORT - University of Cambridge, acessado em julho 13, 2025, https://science.ai.cam.ac.uk/assets/uploads/apsci-annual-report-2024.pdf
104. ADASS2023 :: pretalx, acessado em julho 13, 2025, https://pretalx.com/adass2023/schedule/v/0.9/
105. 2506.15090v1 | PDF | Astronomy | Galaxy - Scribd, acessado em julho 13, 2025, https://www.scribd.com/document/878730165/2506-15090v1
106. 202 4 A nn ual R e p o rt | Jahresb e richt Annual Report Jahresbericht - Heidelberg Institute for Theoretical Studies, acessado em julho 13, 2025,

https://www.h-its.org/wp-content/uploads/2025/06/HITS_Annual_Report_2024_web.pdf
107. ADASS2023 :: pretalx, acessado em julho 13, 2025, https://pretalx.com/adass2023/schedule/
108. Ground Based, Space Based, Infrastructure, Technological Development, and State of the Profession Activities - National Academies, acessado em julho 13, 2025, https://sites.nationalacademies.org/cs/groups/depssite/documents/webpage/deps_193136.pdf
109. What interesting AI projects have you seen in astronomy? - Quora, acessado em julho 13, 2025, https://www.quora.com/What-interesting-AI-projects-have-you-seen-in-astronomy
110. Hurdles for AI for Scientific Discovery - NCBI, acessado em julho 13, 2025, https://www.ncbi.nlm.nih.gov/books/NBK603481/
111. ETHICAL CONSIDERATIONS IN THE USE OF AI FOR ACADEMIC RESEARCH AND SCIENTIFIC DISCOVERY: A NARRATIVE REVIEW, acessado em julho 13, 2025, https://insightsjlss.com/index.php/home/article/download/178/168
112. Science in the age of AI | Royal Society, acessado em julho 13, 2025, https://royalsociety.org/news-resources/projects/science-in-the-age-of-ai/
113. All Eyez on AI: A Roadmap for Science Communication Research in the Age of Artificial Intelligence, acessado em julho 13, 2025, https://jcom.sissa.it/article/pubid/JCOM_2402_2025_Y01/
114. Artificial Intelligence in Science. Challenges, Opportunities and the Future of Research, acessado em julho 13, 2025, https://www.eua.eu/news/member-and-partner-news/artificial-intelligence-in-science-challenges-opportunities-and-the-future-of-research.html
115. Knowledge Integration for Physics-informed Symbolic Regression Using Pre-trained Large Language Models, acessado em julho 13, 2025, https://hj.diva-portal.org/smash/get/diva2:1968832/FULLTEXT01.pdf
116. 4th Nobel Turing Challenge Initiative Workshop | OIST Groups, acessado em julho 13, 2025, https://groups.oist.jp/ja/obu/event/4th-nobel-turing-challenge-initiative-workshop
117. Artificial Intelligence Paradigms for Next-Generation Metal–Organic Framework Research - ACS Publications - American Chemical Society, acessado em julho 13, 2025, https://pubs.acs.org/doi/pdf/10.1021/jacs.5c08214
118. Generative Metascience: A Review of AI as the Next Scientific Instrument and the Emerging Paradigm of Algorithmic Discovery - Preprints.org, acessado em julho 13, 2025, https://www.preprints.org/manuscript/202507.0417/v2/download
119. arXiv:2410.09403v4 [cs.AI] 27 May 2025, acessado em julho 13, 2025, http://arxiv.org/pdf/2410.09403
120. Automating the practice of science: Opportunities, challenges, and implications | Mu Collective, acessado em julho 13, 2025, https://mucollective.northwestern.edu/files/2025-automating-the-practice-of-sci

ence-opportunities-challenges-and-implications.pdf
121. The AI Paradox in Pharma: Hype vs Reality - A Technical Insight, acessado em julho 13, 2025, https://blog.inovia.bio/inovia-bio-insights/the-ai-paradox-in-pharma-separating-hype-from-reality-a-technical-insight
122. AI in science and medicine: A deep dive from the AI Index Report, acessado em julho 13, 2025, https://med.stanford.edu/news/all-news/2025/04/ai-index-report-science-medicine.html
123. AI in my plasma physics research didn't go the way I expected - Hacker News, acessado em julho 13, 2025, https://news.ycombinator.com/item?id=44037941
124. Active Inference AI Systems for Scientific Discovery - arXiv, acessado em julho 13, 2025, https://arxiv.org/html/2506.21329v2
125. From Lab to Clinic: How Artificial Intelligence (AI) Is Reshaping Drug Discovery Timelines and Industry Outcomes - MDPI, acessado em julho 13, 2025, https://www.mdpi.com/1424-8247/18/7/981
126. ISIT 2025: Conference on Intelligent Systems and Information Technologies - CFP, acessado em julho 13, 2025, https://easychair.org/cfp/ISIT2025
127. Artificial Intelligence at the Crossroads of Engineering and Innovation - Scifiniti, acessado em julho 13, 2025, https://scifiniti.com/3006-4163/2/2025.0014
128. The Symbiotic Scientist: Navigating the Evolving AI Landscape in Discovery - Medium, acessado em julho 13, 2025, https://medium.com/@bingcheng00/the-symbiotic-scientist-navigating-the-evolving-ai-landscape-in-discovery-1c1c547b9885
129. (PDF) ETHICAL CONSIDERATIONS IN THE USE OF AI FOR ACADEMIC RESEARCH AND SCIENTIFIC DISCOVERY: A NARRATIVE REVIEW - ResearchGate, acessado em julho 13, 2025, https://www.researchgate.net/publication/391155663_ETHICAL_CONSIDERATIONS_IN_THE_USE_OF_AI_FOR_ACADEMIC_RESEARCH_AND_SCIENTIFIC_DISCOVERY_A_NARRATIVE_REVIEW
130. Large Language Models for History, Philosophy, and Sociology of Science: Interpretive Uses, Methodological Challenges, and Criti - arXiv, acessado em julho 13, 2025, https://www.arxiv.org/pdf/2506.12242
131. TabVI: Leveraging Lightweight Transformer Architectures to Learn Biologically Meaningful Cellular Representations - bioRxiv, acessado em julho 13, 2025, https://www.biorxiv.org/content/10.1101/2025.02.13.637984v1.full.pdf
132. Large Language Model Agent for Modular Task Execution in Drug Discovery - bioRxiv, acessado em julho 13, 2025, https://www.biorxiv.org/content/biorxiv/early/2025/07/05/2025.07.02.662875.full.pdf
133. A Survey of AI for Materials Science: Foundation Models, LLM Agents, Datasets, and Tools, acessado em julho 13, 2025, https://arxiv.org/html/2506.20743v1
134. Search results for `AI epistemology` - PhilArchive, acessado em julho 13, 2025, https://philarchive.org/s/AI%20epistemology

135. Using Mathematics to Explain a Scientific Theory† | Philosophia Mathematica, acessado em julho 13, 2025, https://academic.oup.com/philmat/article/24/2/185/1752454
136. Robert J. Marks II, acessado em julho 13, 2025, https://robertmarks.org/CV/CV-plus.pdf
137. Scientific Discovery Computational Explorations of The Creative Processes by Pat Langley, Herbert A. Simon, Gary L. Bradshaw, Jan M. Zytkow | PDF | Science - Scribd, acessado em julho 13, 2025, https://www.scribd.com/document/728913000/Scientific-Discovery-Computational-Explorations-of-the-Creative-Processes-by-Pat-Langley-Herbert-A-Simon-Gary-L-Bradshaw-Jan-M-Zytkow-z-lib-org
138. Courses - AI FOR SCIENCE MASTERS, acessado em julho 13, 2025, https://ai.aims.ac.za/courses-2024-5
139. Machine Learning – Water Programming: A Collaborative Research Blog, acessado em julho 13, 2025, https://waterprogramming.wordpress.com/tag/machine-learning/
140. SciPy 2025 :: pretalx, acessado em julho 13, 2025, https://cfp.scipy.org/scipy2025/schedule/
141. Current challenges and Inria's engagement, acessado em julho 13, 2025, https://www.inria.fr/sites/default/files/2021-09/Livre_Blanc_IA_MS_20210428%20%281%29.pdf
142. Bayesian Philosophy of Science | Request PDF - ResearchGate, acessado em julho 13, 2025, https://www.researchgate.net/publication/336781037_Bayesian_Philosophy_of_Science
143. Interprétabilité locale post-hoc des modèles de classification "boîtes noires" - LFI | Learning Fuzzy and Intelligent systems, acessado em julho 13, 2025, https://lfi.lip6.fr/wp-content/uploads/2022/10/2020-07-Laugel-Thibault.pdf
144. Beyond Quantity - Research with Subsymbolic AI - SSOAR: Social Science Open Access Repository, acessado em julho 13, 2025, https://www.ssoar.info/ssoar/bitstream/document/90685/1/ssoar-2023-sudmann_et_al-Beyond_Quantity_Research_with_Subsymbolic.pdf
145. Harnessing Artificial Intelligence for Automated Diagnosis - MDPI, acessado em julho 13, 2025, https://www.mdpi.com/2078-2489/15/6/311
146. The Difference between Explainable and Explaining: Requirements, acessado em julho 13, 2025, https://scispace.com/pdf/the-difference-between-explainable-and-explaining-10e4zozhkc.pdf
147. Search results for `relevant similarity` - PhilArchive, acessado em julho 13, 2025, https://philarchive.org/s/relevant%20similarity
148. vol 4 no 2 spring 2024 - International Journal of Law, Ethics, and Technology, acessado em julho 13, 2025, https://ijlet.org/wp-content/uploads/2025/01/IJLET-4.2.pdf
149. Diffusion model - Wikipedia, acessado em julho 13, 2025, https://en.wikipedia.org/wiki/Diffusion_model

150. What are Diffusion Models? | Splunk, acessado em julho 13, 2025, https://www.splunk.com/en_us/blog/learn/diffusion-models.html
151. Diffusion Models: How do They Diffuse? - Towards Data Science, acessado em julho 13, 2025, https://towardsdatascience.com/ai-diffusion-models-how-do-they-diffuse-5ac0fcb4426f/
152. On Proofs and Translation - UC Berkeley EECS, acessado em julho 13, 2025, https://www2.eecs.berkeley.edu/Pubs/TechRpts/2025/EECS-2025-92.pdf
153. Three tiers of computation in transformers and in brain architectures - ResearchGate, acessado em julho 13, 2025, https://www.researchgate.net/publication/389694268_Three_tiers_of_computation_in_transformers_and_in_brain_architectures/fulltext/67ce5775e62c604a0dd6a217/Three-tiers-of-computation-in-transformers-and-in-brain-architectures.pdf
154. Detection or Deception: The Double-Edged Sword of AI in Research Misconduct, acessado em julho 13, 2025, https://www.the-scientist.com/detection-or-deception-the-double-edged-sword-of-ai-in-research-misconduct-72354
155. Can Agentic AI Unlock a Nobel Prize with an Unsolved Mystery? | The AI Journal, acessado em julho 13, 2025, https://aijourn.com/can-agentic-ai-unlock-a-nobel-prize-with-an-unsolved-mystery/
156. What's the Future of AI? - Copyleaks, acessado em julho 13, 2025, https://copyleaks.com/blog/whats-the-future-of-ai
157. AI in science and research, acessado em julho 13, 2025, https://fastdatascience.com/ai-in-research/
158. Is AI leading to a reproducibility crisis in science? - ResearchGate, acessado em julho 13, 2025, https://www.researchgate.net/publication/376254925_Is_AI_leading_to_a_reproducibility_crisis_in_science
159. Artificial intelligence for science: The easy and hard problems - arXiv, acessado em julho 13, 2025, https://arxiv.org/html/2408.14508v1
160. We Need Improved Data Curation and Attribution in AI for Scientific Discovery - arXiv, acessado em julho 13, 2025, https://arxiv.org/html/2504.02486v1
161. A study of 331 retracted chemistry papers found that 69% were due to plagiarism or data manipulation. Only 16% were due to "honest errors." : r/science - Reddit, acessado em julho 13, 2025, https://www.reddit.com/r/science/comments/bucply/a_study_of_331_retracted_chemistry_papers_found/
162. ChatGPT's Expanding Horizons and Transformative Impact Across Domains: A Critical Review of Capabilities, Challenges, and Future Directions - Preprints.org, acessado em julho 13, 2025, https://www.preprints.org/manuscript/202507.0429/v1
163. Capstone-Unified Personal Development Field - Use cases and examples, acessado em julho 13, 2025, https://community.openai.com/t/capstone-unified-personal-development-field/1

253625
164. Fieldwork: Master UX Research in the Age of AI - Eventbrite, acessado em julho 13, 2025, https://www.eventbrite.com/e/fieldwork-master-ux-research-in-the-age-of-ai-tickets-1365035400629
165. 0970-2555 Volume : 53, Issue 5, No.8, May : 2024 UGC CARE Group-1 75 AI DRIVEN DRUG DI - Indian Institution of Industrial Engineering, acessado em julho 13, 2025, http://www.journal-iiie-india.com/1_may_24/14.8.pdf
166. Chatbots and Neural Network - Artificial Intelligence Zone, acessado em julho 13, 2025, https://www.artificialintelligencezone.com/chatbots/neural-network/
167. هوش مصنوعی و آینده پیشرفت علمی: گذار از علم عادی به علم پساعادی, acessado em julho 13, 2025, https://journals.atu.ac.ir/article_16482.html