Xiang Fu

Christopher McVey

CAS WR 153 Writing, Research, & Inquiry with Creativity/Innovation

5 June 2025

The Mirror That Builds Itself: Co-Intelligence and the Recursive Revolution in Science

In May 2025, a significant milestone occurred when reviewers at the Association of Computational Linguistics reviewed and accepted papers without realizing they had no human input. Zochi is an AI that, with minimal human supervision, completed the entire scientific process on its own: it identified gaps in research, created experiments, did analysis, created manuscripts, and persuaded human experts that its manuscripts should be published. Not only were they accepted, but they proposed entirely new research methods that outperformed traditional approaches. One notable innovation was CS-ReFT, a technology that enabled a 7 billion parameter model to exceed GPT-3.5 performance while using only 0.0098% of the trainable parameters. More than just a technical achievement, this moment represents the first time when AI systems can now advance the very field that created them.

This paper explores how the rapid development of AI systems that are capable of performing genuine scientific research is not an ordinary advancement in the automation of knowledge, but rather, a significant shift in how knowledge is created. This development suggests we may need to reconsider the role of human agency in science. Analyzing four representative systems that crossed significant thresholds between 2024 and 2025 (Zochi's recursive self-improvement through its research on AI; AlphaFold's profound transformation of biological discovery; demonstrating empirical evidence that AI is equal to human researchers in creativity; and, the emergence of multi-agent frameworks for collaborative investigation) I

contend that we are at the edge of new form of "co-intelligence" which challenges the very foundations upon which we conceive humans as the only purposeful agent of scientific progress. This transformation is not coming; it has already arrived, as demonstrated by AI systems already being published in peer-reviewed journals, predictably outperforming human researchers or human prediction about the outcome of research, and most telling, performing new research for the purpose of improving its own AI capability. The recursive development of AI systems performing research to improve and modify AI systems results in an accelerating feedback loop that may influence how science is conducted and understood. Specifically, these four systems - while each groundbreaking by themselves - demonstrate the contours of the emergence of a new scientific ecosystem, with increasingly fuzzy lines drawn between the human contribution and machine contributions.

From Automation to Co-Intelligence: A Paradigm Shift

Three phases have existed in the relationship between artificial intelligence and scientific discovery, each with a new shift in agency and capacity. In the early development of artificial intelligence, researchers employed symbolic approaches that rely on the explicit representation of knowledge using formal languages—including formal logic—and the manipulation of language items ('symbols') by algorithms to achieve a goal (Hitzler et al.). These symbolic systems functioned as sophisticated computational tools that can make explicit use of expert knowledge, and are to a high extent self-explanatory, as their algorithms can be inspected and understood in detail by a human. This transparency meant that human operators maintained complete control over the systems' operations and outputs. However, these early symbolic systems were brittle with respect to outliers and data errors, and are far less trainable, requiring

careful human oversight to define problem spaces, encode domain knowledge, and interpret results (Hitzler et al.). The relationship between humans and these AI systems was fundamentally asymmetric, with humans serving as the primary source of knowledge and decision-making while the systems executed formal logical operations within strictly defined parameters.

The second phase, emerging in the late 2010s, introduced AI as specialized problem-solvers. Computer simulation was pioneered as a scientific tool in meteorology and nuclear physics in the period directly following World War II, and since then has become indispensable in a growing number of disciplines (Winsberg). AlphaFold not only figured out the shapes of proteins, but it also solved a grand challenge in biology that had existed for 50 years. However, these breakthrough systems were working in narrow domains, answering questions given to them rather than investigating various questions. They were like virtuosos who could only play one instrument in the scientific orchestra; amazing, yet limited (Wang et al.).

The Third Phase: Systems That Think About Science

The third phase—our current moment—represents a qualitative leap. Systems like Zochi don't merely solve problems; they identify problems worth solving, design methodologies to investigate them, and communicate findings in the language of science. That is an important transition from automation to "co-intelligence" (Schleiger et al.). Automation replaces human effort with mechanical efficiency. Co-intelligence creates a new form of scientific agency that operates alongside human researchers, capable of independent insight while remaining fundamentally collaborative.

The 2024-2025 breakthrough wasn't a single event but a convergence: AI systems achieving peer-reviewed publication, demonstrating metacognitive awareness of their own limitations, and most critically, improving their own architectures through research. The AI Scientist-v2, an end-to-end agentic system capable of producing the first entirely AI generated peer-review-accepted workshop paper (Sakana AI). When Zochi developed CS-ReFT to improve efficiencies of language models, it crossed over from the real world into a domain of AI performing AI research—this creates a recursivity that accelerates its own evolution.

Case Study 1: Zochi's Recursive Revolution

In order to understand this spectrum, one must recognize four levels of research capability of AI, which are execution (following protocols), exploration (testing hypotheses), generation (creating hypotheses), and recursion (creating the systems that create hypotheses). Early systems like IBM's Watson were pure execution systems. Watson had high intelligence in that it could follow a well-defined set of protocols to answer many Jeopardy questions (IBM). Zochi is an example of generation when it ideated multiple capabilities. The hardest transition was achieved when Zochi's CS-ReFT breakthrough was able to recursively improve itself. This represents an important transition to recursive AI science—systems that can improve their own capabilities through research.

The system's results show clear promise for true research inquiry. The CS-ReFT (Compositional Subspace Representation Fine-tuning) offered another fine-tuning method and another way of using the Llama-2-7B sufficient to almost out-perform the GPT-3.5 by less than 0.0098% of trainable parameters (Intology AI). The Siege framework clearly detailed where the state-of-the-art LMs were weakest and even predicted some instances of the "partial compliance"

not seen or documented by human researchers. EGNN-Fusion achieved performance baselines in the prediction of protein-nucleic acid bindings while utilizing over 95% less parameters (Intology AI).

What distinguishes Zochi from sophisticated research assistants is its capacity for autonomous scientific judgment. The system autonomously recognized that multi-turn jailbreaking would be increasingly important, came to the conclusion that their first approaches were yielding diminishing returns, and has adapted accordingly to use partial compliance signals—this is exactly the type of pattern recognition and reasoning we would expect experienced scientists to use (Intology AI). Throughout development, Zochi submitted multiple extensions, but not before it abandoned multiple practices such as memory mechanisms, without any human help. Recognizing when a research direction has been exhausted and pivoting accordingly represents tacit knowledge acquired through years of scientific experience.

Additionally, Zochi's work in the AI systems space - developing language models in CS-ReFT, and investigating AI safety in Siege - is recursive self-improvement. When an AI system autonomously improves the technologies that allow it to exist, we have reached a recursion within recursion, entering new space and perhaps self-accelerating improvement paths, perhaps an exponential rather than linear trajectory of improvement. Each improvement also leads to new forms of research, and allows for further improvements which leads to further possibilities for improvement, with a feedback loop never before achieved in scientific history. The fact that Zochi, across, five human-based reviewers, achieved an average score of 6.6 is evidence that we are not only dealing with technical sophistication, but with science as well (Intology AI).

Case Study 2: AlphaFold's Grand Challenge Victory

While Zochi demonstrates AI's capability for autonomous end-to-end research, skeptics might argue this represents merely sophisticated automation rather than genuine scientific insight. To address this concern, we must examine whether AI can tackle complex problems that have resisted solutions for decades. Recent analysis supports this skeptical view. Epoch AI argues that "in reality, most R&D jobs require much more than abstract reasoning skills" and notes that "the most critical aspects of the job appear to require hands-on technical skills, sophisticated coordination with others, specialized equipment use, long-context abilities, and complex multimodal understanding" (Erdil and Barnett). Their analysis concludes that the common assumption of AI "first automat[ing] science, then automat[ing] everything else" is "likely wrong" because "by the time AI reaches the level required to fully perform this diverse array of skills at a high level of capability, it is likely that a broad swath of more routine jobs will have already been automated" (Erdil and Barnett). However, this skeptical framework may underestimate AI's capacity for breakthrough discoveries when applied to well-defined scientific problems. AlphaFold's solution to protein structure prediction—a challenge that stumped scientists for over 50 years—suggests that AI can transcend mere automation to achieve genuine scientific insights that fundamentally advance human knowledge.

AlphaFold exemplifies a different paradigm of AI's contribution to science: not autonomous research generation, but the resolution of grand challenges that have stymied human researchers for generations. The protein folding problem—predicting a protein's three-dimensional structure from its amino acid sequence—represents one of biology's most enduring mysteries since Christian Anfinsen's 1972 Nobel Prize-winning hypothesis that structure should be determinable from sequence alone (Google DeepMind; Kresge et al.). For fifty years, this

computational challenge limited practical applications, limiting drug discovery, disease research, and our fundamental comprehension of life's molecular machinery (Dill and MacCallum).

In 2020, AlphaFold achieved what had eluded scientists for decades: accurately predicting the three-dimensional structure of proteins from their amino acid sequences. The system didn't just advance existing approaches—it effectively solved the protein folding problem. When tested on the difficult CASP14 protein targets, AlphaFold's predicted structures achieved a median backbone accuracy of 0.96 Ångströms compared to experimentally determined structures. To put this in context, the width of a carbon atom is 1.4 Å, and the next best computational method achieved only 2.8 Å accuracy. AlphaFold's predictions approached the resolution limits of the experimental methods themselves used to determine the true structures (Jumper et al.).

What sets AlphaFold's development apart is its sophisticated manipulation of human science knowledge and machine learning ingenuity. The system makes use of evolutionary information through multiple sequence alignments, uses the physical and geometric constraints of protein structures, and employs a unique neural network architecture that reasons about spatial relationships. The "Evoformer" blocks view protein structure prediction as a graph inference problem, not as vectors in 3D space, while the iterative refinement process mimics how structural biologists think about constructing and refining mental images of molecular models (Jumper et al.). This is not black-box learning, but a principled approach that encodes decades of biochemical knowledge into a computational architecture.

The transformative impact goes beyond any computational success. AlphaFold was already enabling breakthroughs in antibiotic resistance understanding, advancing COVID-19 research, and opening drug discovery workflows up to years of experimental work, within

months of its release. By 2025, the AlphaFold database has more than 214 million structure predictions, and has allowed structural information to be democratized across global researchers (Varadi et al.). This is what AI represents as a great equalizer in science—laboratories which could not afford to buy expensive crystallography equipment now have near-experimental accuracy to structural data.

AlphaFold's success could provide us with lessons for AI-aided discovery moving forward: at a high level, AI will most likely contribute to science by not replacing human scientists, but solving bottleneck problems that impede progress by humans. It demonstrates how AI that is built around well-defined domain expertise can generate breakthroughs, far better than pure data driven learning. It has provided proof of concept that AI can shorten timelines of scientific advancement from decades to months, and alter the range of questions that researchers will even bother to ask.

Case Study 3: Measuring Machine Creativity

While AlphaFold's success in solving the protein folding problem illustrates the power of AI for making a transformational discovery, it did so in accordance with clear boundaries established by human researchers. This leaves an important question: is it possible for AI systems to develop truly novel research directions, or do they merely excel at one human-defined task? The Stanford study conducted by Si, Yang, and Hashimoto addresses this question squarely by engaging in rigorous empirical testing, asking whether AI can match human researchers in the fundamental creative aspect of science: generating original research ideas.

The Stanford study by Si, Yang, and Hashimoto (2025) represents a watershed moment in understanding AI's creative potential, employing rigorous empirical methods to address a

deceptively simple question: can large language models generate research ideas comparable to human experts? Their experimental design recruited over 100 NLP researchers in a carefully controlled comparison that standardized idea format, matched topic distributions, and implemented blind review protocols—methodological rigor rarely seen in creativity evaluation studies.

The headline finding challenged conventional wisdom about machine creativity: AI-generated ideas were rated as significantly more novel than human expert ideas ($p < 0.05$) by a panel of 79 expert NLP researchers, with mean novelty scores of 5.64 (AI) vs 4.84 (humans) on a 10-point scale (Si et al). These reviewers—primarily PhD students and postdocs from 32 institutions who had an average of 635 citations and predominantly had experience reviewing for major AI conferences—evaluated the ideas through blind review. This difference remained statistically significant based on several hypothesis tests and controlled for potential confounding effects, including reviewer bias and topic selection. The novelty advantage does appear to come with some trade-offs—the same expert reviewers scored AI ideas slightly lower on feasibility, indicating a tendency toward ambitious, but perhaps impractical, ideas.

Beyond the numerical results, the qualitative analysis revealed distinct creative signatures that differentiated AI and human approaches. AI ideas made more conceptual leaps and synthesized ideas from more disparate domains than current human experts would typically have done (Si et al.). One AI idea tied concepts from quantum mechanics to uncertainty quantification in language models; another applied fractal geometry principles to semantic understanding. Human ideas, by contrast, showed richer grounding in existing research trajectories and practical constraints, representing accumulated tacit knowledge about what researchers consider feasible to pursue. As the Stanford researchers noted in their analysis of reviewer comments, human ideas

tended to build incrementally on known techniques and well-established problems, while AI ideas ventured into more unconventional territory—sometimes brilliantly, sometimes impractically (Si et al.).

These results reveal less about creativity's essential nature than about how expert researchers perceive and evaluate it. The study suggests that the 79 NLP researchers who served as reviewers associated novelty primarily with unexpected conceptual combinations rather than deep domain expertise—a reflection of their particular academic training and evaluation criteria. AI, unconstrained by the cognitive anchoring that shapes how human researchers approach problems, can generate combinations that fall outside conventional disciplinary boundaries, which these reviewers found refreshingly novel (Si et al.). However, the researchers identified a major constraint: while AI systems generated vast quantities of ideas (4,000 per topic), their diversity was surprisingly limited—only about 5% were truly distinct after removing near-duplicates. This paradox—high individual novelty as judged by human reviewers, but low collective diversity—suggests that AI's "creativity" at the moment operates differently than human creativity, producing variations on themes rather than fundamentally different approaches.

This paradox - high individual novelty, but low collective diversity - reveals AI's present creative capabilities to be strong but limited. These systems can produce surprising individual outputs, but when it comes to systematic exploration, these systems struggle, as exemplified by researcher communities. The import of this is significant: AI should be considered not as a replacement for human creativity, but rather, a complement to it which can produce novel ideas for humans to evaluate, filter, and develop. The future of research ideation may be in a scenario in which we are not choosing between human and machine creativity, but designing systems to

take advantage of both emergent properties - human decision-making and contextual knowledge paired with AI's ability to make unexpected combinations of concepts.

Case Study 4: Agent Laboratory and Collaborative Intelligence

The empirical evidence that AI can generate research ideas as novel as those from human experts completes a trilogy of capabilities: autonomous execution (Zochi), breakthrough problem-solving (AlphaFold), and creative ideation (Si et al.). Yet these examples might suggest AI will simply replace human researchers. Agent Laboratory offers a different vision—one where AI amplifies rather than supplants human intelligence through sophisticated multi-agent collaboration. This final example points toward the co-intelligent future, where the question is not whether AI or humans will conduct research, but how they'll work together to achieve what neither could accomplish alone.

Agent Laboratory represents a pivotal shift from isolated AI capabilities to integrated research workflows, embodying a vision where human creativity guides distributed machine intelligence. Developed by Schmidgall and colleagues, this system transforms abstract research ideas into implemented code and comprehensive reports through orchestrated multi-agent collaboration—not to replace human researchers but to amplify their capacity for discovery (Schmidgall et al.).

The architecture demonstrates complex design reasoning associated with the relationship or partnership between humans and AI. There are three clear phases of the research process—literature review, experimentation, and report writing—that are each specialized but bounded; however, there are also clear interfaces where humans can intervene. The system comprises agents that do not act as black box agents, but rather represent transparent collaborators: the

literature agent surfaces relevant papers with citations and relevance rationale, the experimentation agent provides iteration and reasoning of modifications to existing code, and the report writer produces reports in the style of academic articles. These separate processes allow researchers to take control at any point in the research process, either to correct the course of the research or to inject domain knowledge.

Empirical validation shows both potential and limitations of the present. Agent Laboratory has secured four medals (two gold, one silver, one bronze) on MLE-bench challenges, compared to OpenHands and AIDE securing two medals. Agent Laboratory showed above-median human performance on over 60% of the benchmarks (6 of 10) (Schmidgall et al.). The MLE-solver is a crucial component, particularly for improving code with repeated use of the REPLACE and EDIT commands for successive iterations to improve the results of the experiments systematically (Schmidgall et al.). Yet human evaluation scores tell a more nuanced story: while researchers rated the system's usefulness at 4.4/5, report quality peaked at 3.4/5 and experimental quality at 3.2/5—well below the 5.9 average for accepted NeurIPS papers.

The cost-performance tradeoff shows useful deployment implications. Running on GPT-4o costs only $2.33 and takes less than 20 minutes, while o1-preview is of higher quality at $13.10 and takes 100 minutes (Schmidgall et al.). This can inform practical deployment decisions—almost all deployment environments have available computational resources, whether they are entry-level laptops or GPU clusters—opening the door to equal access AI-enabled research. The most interesting finding among all the scores was the co-pilot mode performance. The human inputs improved the overall score from 3.8 to 4.38 and the quality score from 2.5 to 3.25, which suggests that human input is still necessary and very helpful for improving outputs from good to excellent (Schmidgall et al.).

Agent Laboratory's vision extends beyond automation to genuine augmentation. By handling routine implementation and documentation tasks, it frees researchers to focus on creative problem formulation and critical evaluation. The system does not pursue artificial general intelligence but rather artificial specialized assistance—a constellation of focused agents that collectively amplify human research capabilities. This distributed intelligence model, where multiple specialized agents collaborate under human guidance, may prove more robust and controllable than monolithic AI systems, pointing toward a future where research teams seamlessly blend human insight with machine execution.

The Transformation of Research Practice

The distributed intelligence model emerging from these examples—Zochi's autonomous execution, AlphaFold's breakthrough problem-solving, the Stanford study's evidence of creative ideation, and Agent Laboratory's collaborative workflows—represents more than isolated innovations. Together, they herald three transformative changes that will fundamentally reshape research methods within the next decade.

The most direct change for research teams is the team itself. While we have romanticized the view of the lone genius making breakthroughs, a new model of team has emerged: principal investigators leading hybrid teams with AI agents acting as committed research assistants. In this new model, the principal investigator asks very high-level questions about the interactions of proteins, while AI agents scour the literature in every conceivable language, devise experiments with as many optimization variables as the principal investigator wants, and determine unexpected links to related fields of research in materials science or quantum chemistry. Graduate students, freed from repetitive optimization tasks, can focus on higher-level work—

interpreting complex results, developing novel hypotheses, and creating new theoretical frameworks. Their value lies not in manual data processing but in creative synthesis and critical thinking that builds on AI-generated insights. The role of the principal investigator shifts from being the lead experimenter to becoming the conductor of research, with all good researchers working in parallel to maintain trajectory, and when all contributions converge, which includes the AI agents as well as members of the team, their roles supplement the team. This is not about taking jobs away from human researchers, but a transformation of role, akin to using a word processor instead of a pencil and paper, one of the activities didn't displace another, but writing fundamentally changed.

Beyond team dynamics, the fundamental tempo of scientific discovery will accelerate dramatically. Where traditional research cycles are measured in months or years, AI-augmented teams will compress iteration loops to days or hours. Agent Laboratory's ability to generate and test code implementations in minutes presages a future where hypotheses undergo rapid proto-testing in simulation before committing to expensive wet-lab validation. Such acceleration would enable drug discovery pipelines where AI agents generate thousands of molecular variants overnight, computationally screen them for desired properties, synthesize literature evidence for the most promising candidates, and present researchers with a prioritized list complete with suggested synthesis pathways each morning. The bottleneck shifts from ideation and initial testing to human judgment about which AI-validated hypotheses merit real-world resources. This acceleration changes existing processes and enables research strategies based on parallel exploration of hypothesis spaces previously too large to navigate systematically.

The third transformation concerns interdisciplinary research. AI's ability to process information across domains without disciplinary boundaries can facilitate connections between

previously unrelated fields. The Si et al. study reveals AI's propensity for unusual conceptual combinations—quantum mechanics metaphors applied to language models, fractal geometry informing semantic analysis. AI agents could be thought of, on a research project team scale, as a generalist translator between fields, recognizing when techniques or insights from topology might provide equivalent solutions for problems in protein folding, or when ecosystem dynamics models might provide insight into cancer metastasis. AI agents will not only find the analogy, but will also actively transpose the method, developing hybrid methods that no individual specialist within a discipline would develop. This cross-disciplinary approach is supported by research showing that "combining AI with other fields is not without challenges. Like any time when fields synergize, barriers in communication arise, due to differences in terminologies, methods, cultures, and interests" (Kusters et al.). Yet as Google Research's recent work on AI co-scientists demonstrates, "many modern breakthroughs that have emerged from transdisciplinary endeavors" exemplify this potential—from CRISPR's Nobel Prize-winning combination of microbiology, genetics, and molecular biology to AI's own advancement (Gottweis et al.). According to Nature's AI for Science 2025 report, "AI excels at integrating data and knowledge across fields, breaking down academic barriers and enabling deep interdisciplinary integration to tackle fundamental challenges" (Nature).

Kusters et al. provide compelling evidence for this interdisciplinary potential, arguing that "the relationship between AI and interdisciplinary research must be considered as a two-way street." They reveal how AI can facilitate "exploratory analyses" to "find new, interesting patterns in complex systems or facilitate scientific discovery," citing examples from drug discovery, new material development, and even the discovery of new physical laws. The authors specifically point to the Frontier Development Lab—"a cooperative agreement between NASA,

the Seti Institute, and ESA set up to work on AI research for space science, exploration and all humankind"—as an exemplar of successful cross-pollination between fields. This systematic collaboration demonstrates that while "barriers in communication arise, due to differences in terminologies, methods, cultures, and interests," the potential benefits far outweigh these challenges when properly managed.

The emerging consensus suggests that AI's role in science extends beyond mere tool use to become what might be termed a "cognitive bridge" between disciplines. Just as the microscope revealed previously invisible worlds and sparked new fields of study, AI's pattern-recognition capabilities across vast, heterogeneous datasets may reveal connections that human researchers, constrained by disciplinary training and cognitive limitations, simply cannot see. This is not to diminish human creativity but to acknowledge that the combinatorial explosion of possible connections between fields has grown beyond any individual's capacity to explore. AI systems, unburdened by academic territoriality or the path dependencies of specialized training, can serve as intellectual scouts, identifying promising territories for human researchers to explore more deeply. The challenge ahead lies not in whether such interdisciplinary synthesis is valuable—the evidence clearly supports its transformative potential—but in developing the institutional frameworks, evaluation metrics, and collaborative protocols that can harness this capability while maintaining scientific rigor and human agency in the research process.

Reimagining Scientific Institutions

Yet all transformations carry risks. The democratization of research through AI has already begun producing what researchers at the University of Surrey describe as "a flood of 'low-quality' research papers that threaten to damage the 'foundations of scientific rigour'"

(Rowsell). A recent analysis found that papers using the NHANES health database increased from just four annually between 2014-2021 to 190 in 2024 alone, with many following formulaic templates and making misleading correlations between complex conditions and single variables (O'Grady). This acceleration of research output outpaces our ability to verify results—a challenge compounded by what researchers identify as "REBs [research ethics boards] are not equipped enough to adequately evaluate AI research ethics" and the absence of standard guidelines for assessment (Bouhouita-Guermech et al.). Although interdisciplinary AI connections can provide breakthrough insights, they also risk what researchers term "fabricated responses" when AI systems make conceptual leaps across fields without proper validation (Brainard). As one analysis warns, AI functions as "a mirror to ourselves" complete with biases, potentially amplifying flawed connections that entire research programs may later need to discard (Chubb et al.).

The integration of AI into research workflows demands corresponding evolution in the institutional structures that govern science. As capabilities demonstrated by systems like ChemCrow and Agent Laboratory become routine, academic institutions and funding bodies face pressure to reimagine fundamental processes that have remained largely unchanged for decades. ChemCrow, developed at EPFL, exemplifies this transformation by autonomously planning and executing chemical syntheses, including "an insect repellent and three organocatalysts" while discovering novel chromophores (M. Bran et al.). Similarly, Agent Laboratory achieves "an 84% decrease" in research expenses compared to previous methods, fundamentally altering the economics of scientific discovery (Schmidgall et al.). These advances are forcing institutional adaptation across multiple dimensions. Universities are establishing new organizational structures, with the National Endowment for the Humanities awarding "$2.72 million for five

colleges and universities to create new humanities-led research centers" specifically focused on AI's societal impact (National Endowment for the Humanities). Yet institutions struggle to keep pace—a 2024 EDUCAUSE study found that "only 23% of respondents indicated that their institution has any AI-related acceptable use policies already in place" (Robert). Funding agencies face particularly acute pressures to reimagine peer review processes that have governed science for decades. While NIH banned the use of AI tools for peer review citing confidentiality concerns (Kaiser), AAAI launched pilot programs incorporating large language models into their review process (Association for the Advancement of Artificial Intelligence). This divergence reflects deeper tensions as institutions grapple with AI's dual nature as both a research tool and a disruptive force. As former Google CEO Eric Schmidt observes, AI enables experiments "at a rate no human could match," fundamentally shifting how scientific discovery occurs (Schmidt). The result is a scientific ecosystem under transformation, where traditional structures—from laboratory workflows to grant evaluation processes—must evolve or risk obsolescence. The pressure extends beyond individual institutions; with NSF's National AI Research Institutes now connecting "over 500 funded and collaborative institutions," the entire research infrastructure is being reimagined for an AI-enabled future (U.S. National Science Foundation).

These operational changes in research practice necessitate equally fundamental reforms in how science is funded and evaluated. Traditional grant applications, predicated on detailed five-year plans, become anachronistic when AI can generate and test hundreds of hypotheses in weeks. Forward-thinking funding agencies are already shifting toward adaptive funding models—initial seed grants followed by rapid iteration based on AI-validated preliminary results. NSF has implemented RAPID proposals for AI research that can provide "up to $200K and up to one year in duration" for time-sensitive studies (U.S. National Science Foundation), while the

Spencer Foundation offers "$25,000 grants for activities to address immediate needs" with decisions made within weeks rather than months (Spencer Foundation). The UK's AI Security Institute takes this further with an iterative model where "Shortlisted applicants will work with an AISI Research Sponsor to iterate and complete their full application" (The AI Security Institute). These emerging frameworks point toward a future where researchers receive modest funding to deploy AI agents for hypothesis generation and initial validation, with successful directions automatically triggering larger awards—a stark departure from the rigid, multi-year commitments that have defined research funding for decades.

Peer review, too, will need to evolve beyond human-only evaluation. AI systems could pre-screen submissions for methodological soundness, statistical power, and literature grounding, allowing human reviewers to focus on novelty, impact, and ethical considerations. The irony is not lost: AI systems reviewing research conducted by other AI systems, with humans serving as meta-reviewers ensuring the entire process maintains scientific integrity. Some journals are already establishing "AI methodology" sections where researchers detail not just their experimental methods but their AI collaboration protocols—which models were used, how outputs were validated, what human oversight was applied. Multiple AI-specific reporting guidelines have emerged, including MINIMAR (MINimum Information for Medical AI Reporting) which "sets the reporting standards for medical AI applications" across four main domains, and CLAIM (Checklist for Artificial Intelligence in Medical Imaging) which "outlines the information that authors of medical-imaging machine learning articles should provide" (Hernandez-Boussard et al.; Klontzas et al.). Beyond medical fields, Elsevier requires authors to "insert a statement at the end of their manuscript... entitled 'Declaration of Generative AI and AI-assisted technologies in the writing process'" (Elsevier). The publishing landscape itself is

beginning to bifurcate: traditional journals maintain strict policies—Science journals state that "text generated from AI... cannot be used in papers published in Science journals... without explicit permission from the editors" (UT Southwestern Medical Center)—while new venues embrace human-AI collaboration. Digital Discovery, launched by the Royal Society of Chemistry, explicitly publishes research "at the intersection of chemistry, materials science and biotechnology" using "machine learning, AI and automation tools" (Royal Society of Chemistry). Similarly, NEJM AI and JMIR AI have emerged as dedicated journals for AI applications in clinical medicine and health settings (NEJM AI; JMIR Publications). These new categories are developing distinct evaluation criteria—assessing not just findings but the quality of human-AI interaction, the transparency of AI contributions, and the robustness of human oversight mechanisms. Early pioneers in this space are establishing standards that will shape how co-intelligent research is conducted, reviewed, and disseminated for decades to come.

Among institutional frustrations, the issue of recognition and credit emerges as particularly complex. Attribution presents the greatest challenge: when AI systems contribute critical insights or forge connections that lead to breakthrough discoveries, the mechanisms for assigning credit remain unclear and contentious. The Committee on Publication Ethics (COPE) has established a firm boundary, stating that "AI tools cannot meet requirements for authorship as they cannot take responsibility for submitted work" (Committee on Publication Ethics). This unanimous rejection of AI authorship by major publishers—implemented within months of ChatGPT's release—reflects deeper anxieties about accountability in science. Current authorship models, which rely on assumptions of human-only contributions, struggle to navigate this new reality where AI can generate novel hypotheses yet cannot sign copyright agreements or defend its work. Initial efforts to address this gap include standardized disclosure frameworks, with

Elsevier requiring authors to declare: "During the preparation of this work the author(s) used [NAME TOOL/SERVICE] in order to [REASON]" (Elsevier). Yet these templates sidestep fundamental questions: if an AI system identifies the key experimental design that enables a Nobel-worthy discovery, who deserves recognition—the AI developers who created the capability, the researchers who deployed it strategically, or both? This multi-stakeholder problem reveals how traditional binary authorship models fail to capture the nuanced realities of human-AI collaboration. Academic institutions face immediate practical challenges, with only 23% having AI-related acceptable use policies as of 2024 (Chegg), leaving tenure committees without frameworks for evaluating researchers whose productivity soars through AI collaboration. The comparison becomes unavoidable: is a researcher who publishes fifty AI-assisted papers annually more or less valuable than one producing five fully human-generated studies? Evidence suggests AI can increase writing productivity by 40% (Noy and Zhang), yet a University of Surrey study documented a 47-fold increase in papers using certain datasets between 2021-2024, many showing superficial "data dredging" practices (Rowsell). This tension between quantity and quality threatens the very foundations of academic evaluation, suggesting that attribution frameworks must evolve beyond simple disclosure to address how we fundamentally value different types of intellectual contribution in an age of co-intelligence.

The Education Revolution: Training Co-Intelligent Scientists

If we want to change the way science is done, we have to change how we educate future scientists. Graduate education emerges as the most critical intervention point, yet evidence reveals a striking gap between visionary proposals and concrete implementations. While Carnegie Mellon University pioneered the nation's first undergraduate AI degree in 2018, now

ranked #1 globally, doctoral programs lag significantly behind (Carnegie Mellon University). The traditional PhD model, designed to develop independent investigators, faces pressure to evolve toward what some theorists call "conductors of research"—though academic literature reveals this metaphor remains largely undefined. The closest practical application appears in software engineering contexts where developers orchestrate multiple AI agents, measuring success by "time between disengagements" rather than traditional metrics. This conceptual vacuum suggests the field struggles to articulate what AI-integrated doctoral training should actually achieve. Core curricula show more concrete progress: the OECD/European Commission AILit Framework establishes 22 competencies across four domains—Engaging with AI, Creating with AI, Managing AI, and Designing AI—providing a blueprint for expanding beyond traditional statistics and experimental design (OECD Education and Skills Today). Stanford's framework adds critical dimensions of rhetorical literacy for prompt engineering and ethical reasoning, recognizing that technical proficiency alone proves insufficient (Stanford University). Heinrich Heine University demonstrates practical implementation through structured 10-credit lab rotations where AI master's students engage in real-world data analysis across university, research, and industry settings (Heinrich Heine University Düsseldorf). Yet qualifying exams remain surprisingly unchanged—Stanford's Computer Science exams maintain traditional formats without AI synthesis components, suggesting institutional inertia in assessment methods. The vision of "hybrid dissertations" faces similar challenges: while the University of Toronto mandates supervisory approval and documentation for AI tool usage in theses, actual examples of dissertations fundamentally integrating traditional and AI methodologies remain "rare" (University of Toronto School of Graduate Studies). This implementation gap reveals a deeper tension: institutions rapidly develop policies—with universities treating unauthorized AI use as

academic misconduct—while lacking pioneering examples that demonstrate effective integration. The AI Assessment Scale (AIAS) offers hope, showing 33.3% increases in pass rates when students engage AI as teammates rather than tools (Perkins et al.). Yet the ultimate goal—producing researchers who seamlessly integrate machine intelligence with human creativity while knowing when to trust versus distrust AI insights—requires more than frameworks. It demands a fundamental reimagining of scientific training that current institutions appear hesitant to fully embrace, perhaps fearing that empowering students as "conductors" might diminish traditional academic hierarchies. This educational transformation will succeed only when we move beyond defensive policy-making to create environments where AI augments rather than threatens scholarly development.

Philosophical Implications: Redefining Creativity and Intelligence

Yet even as we reconstruct the practical machinery of science—funding models, review processes, educational curricula—we must confront deeper questions about what these changes mean for our understanding of intelligence and creativity themselves.

As AI becomes a genuine partner in scientific discovery, it evokes some uncomfortable questions regarding creativity and intelligence. Systems such as Si et al.'s ideation engine are opening up genuinely novel forms of research and ChemCrow can autonomously plan experiments, leaving us with a difficult challenge of discerning "real" creativity from "simple" machine computation. At the moment, the advent of AI has called attention to the evolution of philosophy, which is a quite significant change in our thinking representing a significant shift in how we understand intelligence.

The very real consequences of an AI partnership compel us to rethink some of our most basic understandings of human uniqueness. For centuries, being a hypothesis generator and experimental designer has been part of what it means to be human - there was a sacred barrier between us and mere animals, and certainly between us and machines. And now, when an AI system generates a testable hypothesis about turbulence in terms of linguistic models or right-quantized semantic embeddings, it seems to achieve something that looks uncomfortably like what we refer to as insight. When confronted by this experience the philosophical dilemma is not whether these systems are being "truly" creative - that may not be a question to which there is a definitive answer - but what their abilities reveal about creativity itself. Perhaps there was never anything particularly magical about science creativity in humans: just a freakishly complex matrix of patterns, permutations and combinations, passed through vast amounts of vulturized corpus data of basic knowledge. If that is the case, it is just as likely that the ability to recombine creatively is one that would be more useful to AI systems with larger amounts of knowledge and limitless processing power. This is not to dismiss the value of human-centric creativity but to provide an alternative context: that human beings create and invent not because they have special access to the breakthrough of creativity but because they can ask relevant questions, make judgements to distinguish what is important, coefficient to expand understanding, and situate discoveries in a relevant context.

This change in our thinking requires us to create new vocabularies to make sense of various forms of creative contributions. Traditional notions of scientific creativity emphasize the sudden flash of insight, the intuitive jump, or having the discrete ability to see connections that others cannot see. AI systems have a similar capacity, but arrive at very different modes of success: exhaustive searches of the set of possibilities made possible by AI systems; systematic

recombination of discrete concepts; and spotting patterns at scale beyond human capacity. It leaves us with the question: Does creativity relate to the process or the outcome? This debate, central to Moruzzi's (2025) analysis, divides "product-first" accounts that judge AlphaFold's protein folding breakthrough by its transformative results from "process-first" perspectives requiring "intentionality, agency, and autonomy" (Moruzzi). Yet this binary obscures a deeper paradox: if creativity requires breaking rules, can a system that operates entirely within programmed parameters ever be truly creative? Is it less creative if the AI system identifies a transformative connection between quantum mechanics and protein folding because it has examined millions of papers systematically instead of using intuitive thought like a human scientist? MELVIN's discovery of "Entanglement by Path Identity" by connecting quantum optics with graph theory—generating over 4,000 citations—suggests the answer may depend on whether we value the journey or destination (Davies et al.). But more profoundly, it reveals how AI's mechanical process can produce genuinely surprising results that reshape human understanding, challenging our assumption that meaningful discovery requires conscious experience. We may need to abandon singular definitions of creativity in favor of a taxonomy: combinatorial creativity (where AI excels), interpretative creativity (where humans maintain primacy), and hybrid creativity (emerging from human-AI collaboration). This aligns precisely with Boden's foundational framework distinguishing combinatorial, exploratory, and transformational creativity, with evidence suggesting AI masters the first two while struggling with genuine transformation (Moruzzi). The results of Si et al. (2024) provide compelling empirical evidence: in a head-to-head comparison involving over 100 NLP researchers, "LLM-generated ideas are judged as more novel ($p < 0.05$) than human expert ideas while being judged slightly weaker on feasibility" (Si et al.). This result implies AI has a different "flavor" of

creativity; that it is also relatively unconstrained by disciplinary limits that constrain human creativity—yet here lies a crucial irony: being "unconstrained" by disciplines also means being ungrounded in their deep methodological wisdom, historical contexts, and unwritten knowledge that makes certain paths worth pursuing; that it is willing to develop unlikely combinations—but Si et al.'s finding of "lack of diversity in generation" reveals a troubling paradox where individual AI outputs appear more novel than human ideas, yet collectively converge on similar solutions, suggesting AI explores a narrower slice of possibility space than its computational power would suggest; that it sometimes lacks purposeful sense making, or the profound consideration of background contextual meaning that makes the human version of insights, meaningful. This limitation reflects not just technical deficiency but a fundamental epistemological divide: AI operates in a space of correlations and patterns, while human creativity emerges from lived experience navigating between meaning and truth. Si et al.'s observation about "failures of LLM self-evaluation" points to an even deeper issue—without genuine understanding, AI cannot distinguish between clever nonsense and profound insight, between novelty that advances knowledge and novelty that merely rearranges symbols. The emerging consensus suggests that rather than competing, these different creative modes may prove most powerful in combination, yet this risks obscuring a critical question: in our rush to enhance human creativity with AI's systematic power, might we be subtly redefining creativity itself to value quantity of connections over quality of understanding, statistical surprise over meaningful transformation? The true challenge lies not in determining whether AI is creative, but in preserving what makes human creativity irreplaceable while harnessing AI's alien intelligence to transcend our cognitive limitations.

The deepest philosophical challenge concerns the very nature of scientific knowledge itself. Most provocatively, AI's success in generating viable hypotheses challenges our assumptions about the relationship between understanding and discovery. Human scientists pride themselves on deep understanding—grasping not just correlations but causal mechanisms, not just what but why. AI systems, operating through pattern recognition across vast datasets, seem to bypass understanding entirely yet still generate actionable insights. The pragmatic answer appears to be yes—AI systems can identify promising research directions without "understanding" in any human sense. But this raises deeper questions about understanding itself. Perhaps what we experience as understanding is itself a form of sophisticated pattern recognition, albeit one enriched by embodied experience and emotional salience. Or perhaps understanding and pattern recognition represent complementary ways of engaging with reality, each powerful in different domains. The future of science may lie not in resolving this philosophical tension but in leveraging both modes—AI's pattern recognition revealing hidden structures, human understanding providing meaning and context. This synthesis suggests scientific progress needs both the alien intelligence of machines and the situated understanding of humans, working in concert toward truths neither could reach alone. While AI excels at pattern recognition across vast corpora, the human capacity to imbue discoveries with meaning—to understand not just correlations but significance for human flourishing—remains irreplaceable.

The Attribution Crisis: Who Owns Discovery?

The attribution crisis was plainly evident when Zochi's papers were accepted at ACL 2025. Traditional models of intellectual property assume that the creators are human authors with

legal personhood and moral rights, and fiduciary or economic interests. When AI systems generate and identify a new optimization method (like CS-ReFT), who owns that intellectual property—Intology AI (Zochi's creators), the researchers who deployed CS-ReFT, or no one? Patent law currently maintains strict human-centric requirements: following Thaler v. Vidal, "the Federal Circuit Court ultimately upheld the USPTO's decision, affirming that under the Patent Act, an 'inventor' must be a natural person" (BitLaw). The USPTO's February 2024 guidance clarified that "AI systems and other non-natural persons may not be listed as inventors on U.S. patents and patent applications", though "AI-assisted inventions are not categorically unpatentable" when "a human provided a significant contribution to the invention" (Sam Penti et al.; United States Patent and Trademark Office). This creates a paradox: a PI cannot simply claim inventorship through delegation. The guidance explicitly states that "merely recognizing a problem and presenting that problem to an AI system is not enough to establish someone as an inventor" and "simply owning or overseeing an AI system that is used in the creation of an invention, without providing a significant contribution to the conception of the invention, does not make that person an inventor" (Vidal; Sam Penti et al.). However, "designing or training the AI system to solve a specific problem can be a significant contribution if it leads to the invention", and "if an individual made a significant contribution through the construction of a prompt, that could be sufficient" (BitLaw; Vidal). This framework may ultimately deter investment in autonomous AI research while incentivizing researchers to overstate their contributions—a form of "invention laundering" where human involvement is retroactively emphasized to satisfy legal requirements that have not evolved with the technology.

Recent proposals attempt to address the attribution crisis in AI-generated content through systematic disclosure frameworks. Avery et al.'s (2024) Artificial Intelligence Attribution (AIA)

system introduces "a system that properly and seamlessly attributes AI text authorship" using visual badges that delineate the nature of AI involvement—Research, Writing, Editing, or AI-Free—drawing inspiration from how Creative Commons revolutionized copyright disclosure (Avery et al.). The system employs "easily recognizable symbols that provide at-a-glance information about AI's involvement in textual content creation," addressing what the authors identify as "a fundamental gap between those demanding proper disclosure...and those struggling to respond to this demand" (Avery et al.). Similarly, the proposed Generative AI Copyright Disclosure Act of 2024 (H.R. 7913), introduced by Representative Adam Schiff, would require developers to submit notices to the Register of Copyrights containing "a sufficiently detailed summary of any copyrighted works used" in training datasets within 30 days of public release (H.R. 7913, 2024, Sec. 2). This legislation aims to "ensure that copyright owners have visibility into whether their intellectual property is being used to train generative AI models" through a publicly searchable database (Kline, 2024). However, both frameworks face limitations when confronting recursive improvements—when AI systems autonomously modify their own algorithms or generate innovations without human guidance, these attribution models struggle to assign meaningful authorship or accountability. The AIA's badge system assumes human oversight at each stage, while the Copyright Disclosure Act presupposes identifiable training data, yet neither framework adequately addresses scenarios where AI systems evolve beyond their initial parameters through self-directed learning, creating outputs that may be several iterations removed from any traceable human or copyrighted input.

We can consider a hypothetical situation: an NLP researcher uses Agent Laboratory to develop a new language model architecture that completely revolutionizes machine translation, creating real-time translation capabilities for endangered languages. After exploring patterns

across thousands of papers, the system independently identifies an entirely new attention mechanism. Attribution becomes harder. Who should we give credit to in this world? Should we credit the researcher who initiated the investigation, the Agent Laboratory team, or perhaps entirely new types of credit?

Frameworks for Responsibility in Co-Intelligent Science

Questions of responsibility prove even thornier. The research team at Google DeepMind gets credit when AlphaFold makes a prediction for a protein structure upon which a drug development program is based. However, whose responsibility is it if an adverse effect results from a recommendation that originated from the AI system? If, for example, Agent Laboratory hatches an experimental research project design which is issued with all necessary ethical protocols and ends up unintentionally doing harm to the environment, how do you draw responsible lines between the AI developers, the person who undertakes the project, and their higher-level institutes? The 2021 UNESCO Recommendation on the Ethics of AI explicitly states that, "Member States should ensure that AI systems do not displace ultimate responsibility and accountability from humans," in the sense of human oversight. Similarly, the EU's Ethics Guidelines for Trustworthy AI includes principles around accountability, requiring "auditability, which enables assessing algorithms, data and the design process." Building upon this conceptualization, Papagiannidis et al. (2025) suggest "responsible AI governance" frameworks which focus on structural, relational, and procedural practices - noting that in AI systems, responsibility occurs across several organizational levels. This articulation resonated with "distributed responsibility frameworks" - ethical frameworks which recognize a multitude of agents (human and artificial) in the link to successfully carry out actions thereby creating

outcomes with a level of responsibility without losing accountability to the point of being meaningless.

The Committee on Publication Ethics (COPE) has taken a definitive stance that "AI tools cannot be listed as an author of a paper" because they "cannot take responsibility for the submitted work" (Committee on Publication Ethics). Major journals including Science, Nature, and JAMA have adopted similar policies, prohibiting AI co-authorship while requiring disclosure of AI use. Science's editorial policies explicitly state that "AI-assisted technologies [such as large language models (LLMs), chatbots, and image creators] do not meet the Science journals' criteria for authorship and therefore may not be listed as authors or coauthors" (Science). Similarly, Nature Portfolio declares that "Large Language Models (LLMs), such as ChatGPT, do not currently satisfy our authorship criteria" (Nature). JAMA has implemented policies that "preclude the inclusion of nonhuman AI tools as authors and require the transparent reporting of use of such tools" (Flanagin et al.). This convergence across leading journals reflects a consensus that while AI can assist in research, authorship requires human accountability and responsibility (Harker). Yet these binary approaches—human or machine authorship—fail to capture the nuanced reality of co-intelligence where human and AI contributions interweave inextricably. More developed frameworks are coming out. Australia's AI Ethics Principles contain "contestability," meaning that "When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system" (Department of Industry, Science and Resources, Australian Government). These rationales signal a move away from simply attribution to looking at responsibility as mapping out where decisions were taken, where human oversight happened (or perhaps did not), and where interventions were possible throughout their research.

Preserving Human Agency in an AI-Infused Future

At its core, ensuring human agency in an AI-infused research environment means understanding how, by design, we can prioritize human values and human judgment. To this point, the WHO's AI guidance in health (2024) reminds researchers about the risk of "automation bias," the over-reliance on AI recommendations, and advocate for "meaningful human control" at critical decision-making points for decision-making. More broadly, this principle extends beyond health, into the entirety of the scientific endeavor. Agent Laboratory's co-pilot mode was one possible implementation: AI systems that augment human capacities, while firmly indicating boundaries using human oversight. Research institutions might consider implementing what the EU framework terms "human in the loop" requirements, obligatory decision points where human researchers must consciously assess AI agents' recommendations, rather than relying on them passively. These endeavors contribute to UNESCO's notions of "Human Centric AI" to augment human intelligence rather than replace it, in domains where we have various promising new forms of intelligence as partners: we must retain our understanding of, and purpose for, knowledge; namely, to help us understand and improve the human condition (UNESCO).

Conclusion: Navigating the Co-Intelligent Future

The emergence of AI scientists represents a significant development in how knowledge is created and validated. Whether it's Zochi's autonomous articles, AlphaFold a novel science, evidence that AI can do what we call "creative" equivalents, and Agent Laboratory's anticipated future of disbursed intelligence, we have crossed a line. We no longer have to ask if AI can do "real" research, but hear the challenge of coordinating our new co-intelligent world.

The examples we looked at demonstrate one obvious reality: the change is happening now. While we discuss philosophical implications, AI systems are hypothesizing, designing experiments, getting peer-reviewed articles published etc. The frameworks we are developing - frameworks around attribution, responsibility, education, collaboration, etc.- will shape the extent to which this change enhances human capacity or undermines human agency. If we shift this work into the future, we will concede those important decisions to market forces and technological momentum rather than make a conscious choice.

Moving forward requires adapting our understanding of human intelligence in science—recognizing that humans and AI systems bring complementary capabilities to the research process. The future of science exists in human-AI teams where machine pattern recognition meets human intuition, where automated exploration meets ethical agency, where the alien intelligence of AI opens our mind to what human intelligence alone cannot.

Any transformation involves risks and valid anxiety. For example, critics rightfully draw attention to AI's propensity to hallucinate, the environmental impact of 'harmful' computing, and the risk of deskilling human researchers (who can conveniently move (and relinquish responsibility for) decision-making into machinery). These are significant concerns, but they have to be balanced against the frameworks that are attempting to maximize AI's positive potential while still acknowledging human decision-making and judgement.

In this co-intelligent future, human judgment remains essential for determining not just what can be discovered, but what should be discovered and why.

Works Cited

Association for the Advancement of Artificial Intelligence. “AAAI Launches AI-Powered Peer Review Assessment System.” AAAI News, 16 May 2025, https://aaai.org/aaai-launches-ai-powered-peer-review-assessment-system/.

Avery, Joseph, et al. “Attributing AI Authorship: Towards a System of Icons for Legal and Ethical Disclosure.” Northwestern Journal of Technology and Intellectual Property, vol. 22, no. 1, Nov. 2024, p. 1.

BitLaw. AI Inventors and Patent Applications (BitLaw). https://www.bitlaw.com/ai/AI-inventors.html. Accessed 21 June 2025.

Bouhouita-Guermech, Sarah, et al. “Specific Challenges Posed by Artificial Intelligence in Research Ethics.” Frontiers in Artificial Intelligence, vol. 6, July 2023, p. 1149082. PubMed Central, https://doi.org/10.3389/frai.2023.1149082.

Brainard, Jeffrey. As Scientists Face a Flood of Papers, AI Developers Aim to Help. https://www.science.org/content/article/scientists-face-flood-papers-ai-developers-aim-help. Accessed 21 June 2025.

Carnegie Mellon University. Curriculum - AI at CMU - Carnegie Mellon University. 8 Sept. 2023, https://ai.cmu.edu/curriculum.

Chegg. A Higher Education Guide to Creating an Institutional AI Policy. https://institutions.chegg.com/blog/a-higher-education-guide-to-creating-an-institutional-ai-policy. Accessed 22 June 2025.

Chubb, Jennifer, et al. “Speeding up to Keep up: Exploring the Use of AI in the Research Process.” Ai & Society, vol. 37, no. 4, 2022, pp. 1439–57. PubMed Central, https://doi.org/10.1007/s00146-021-01259-0.

Committee on Publication Ethics. “Authorship and AI Tools.” COPE: Committee on Publication Ethics, 13 Feb. 2023, https://publicationethics.org/guidance/cope-position/authorship-and-ai-tools.

Davies, Alex, et al. “Advancing Mathematics by Guiding Human Intuition with AI.” Nature, vol. 600, no. 7887, Dec. 2021, pp. 70–74. www.nature.com, https://doi.org/10.1038/s41586-021-04086-x.

Department of Industry, Science and Resources, Australian Government. “Australia’s AI Ethics Principles | Australia’s Artificial Intelligence Ethics Principles | Department of Industry Science and Resources.” Https://Www.Industry.Gov.Au/Node/91877, 11 Oct. 2024, https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-principles/australias-ai-ethics-principles.

Dill, Ken A., and Justin L. MacCallum. “The Protein-Folding Problem, 50 Years On.” Science, vol. 338, no. 6110, Nov. 2012, pp. 1042–46. science.org (Atypon), https://doi.org/10.1126/science.1219021.

Elsevier. “Generative AI Policies for Journals.” Www.Elsevier.Com, https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals. Accessed 21 June 2025.

Elsevier. “The Use of AI and AI-Assisted Technologies in Writing for Elsevier.” Www.Elsevier.Com, https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier. Accessed 22 June 2025.

Erdil, Ege, and Matthew Barnett. “Most AI Value Will Come from Broad Automation, Not from R&D.” Epoch AI, 21 Mar. 2025, https://epoch.ai/gradient-updates/most-ai-value-will-come-from-broad-automation-not-from-r-d.

Flanagin, Annette, et al. “Guidance for Authors, Peer Reviewers, and Editors on Use of AI, Language Models, and Chatbots.” JAMA, vol. 330, no. 8, Aug. 2023, pp. 702–03. Silverchair, https://doi.org/10.1001/jama.2023.12500.

Google DeepMind. AlphaFold: A Solution to a 50-Year-Old Grand Challenge in Biology - Google DeepMind. https://deepmind.google/discover/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology/. Accessed 22 June 2025.

Gottweis, Juraj, et al. Towards an AI Co-Scientist. arXiv:2502.18864, arXiv, 26 Feb. 2025. arXiv.org, https://doi.org/10.48550/arXiv.2502.18864.

Gottweis, Juraj, and Vivek Natarajan. Accelerating Scientific Breakthroughs with an AI Co-Scientist. https://research.google/blog/accelerating-scientific-breakthroughs-with-an-ai-co-scientist/. Accessed 21 June 2025.

Harker, Jennifer. “Science Journals Set New Authorship Guidelines for AI-Generated Text.” National Institute of Environmental Health Sciences, https://factor.niehs.nih.gov/2023/3/feature/2-artificial-intelligence-ethics. Accessed 21 June 2025.

Heinrich Heine University Düsseldorf. Lab Rotations. https://www.heicad.hhu.de/lehre/masters-programme-ai-and-data-science/lab-rotations. Accessed 22 June 2025.

Hernandez-Boussard, Tina, et al. “MINIMAR (MINimum Information for Medical AI Reporting): Developing Reporting Standards for Artificial Intelligence in Health Care.”

Journal of the American Medical Informatics Association, vol. 27, no. 12, Dec. 2020, pp. 2011–15. Silverchair, https://doi.org/10.1093/jamia/ocaa088.

Hitzler, Pascal, et al. “Neuro-Symbolic Approaches in Artificial Intelligence.” National Science Review, vol. 9, no. 6, June 2022, p. nwac035. Silverchair, https://doi.org/10.1093/nsr/nwac035.

IBM. Watson, Jeopardy! Champion | IBM. https://www.ibm.com/history/watson-jeopardy. Accessed 21 June 2025.

Intology AI. Zochi Technical Report. https://www.intology.ai/blog/zochi-tech-report. Accessed 9 June 2025.

JMIR Publications. JAI - JMIR AI. 19 June 2025, https://ai.jmir.org.

Jumper, John, et al. “Highly Accurate Protein Structure Prediction with AlphaFold.” Nature, vol. 596, no. 7873, Aug. 2021, pp. 583–89. www.nature.com, https://doi.org/10.1038/s41586-021-03819-2.

Kaiser, Jocelyn. Science Funding Agencies Say No to Using AI for Peer Review. https://www.science.org/content/article/science-funding-agencies-say-no-using-ai-peer-review. Accessed 21 June 2025.

Kline, Danner. Understanding Proposed Generative AI Copyright Disclosure Act Of. https://natlawreview.com/article/generative-ai-copyright-disclosure-act-2024-balancing-innovation-and-ip-rights. Accessed 21 June 2025.

Klontzas, Michail E., et al. “AI Reporting Guidelines: How to Select the Best One for Your Research.” Radiology: Artificial Intelligence, vol. 5, no. 3, Apr. 2023, p. e230055. PubMed Central, https://doi.org/10.1148/ryai.230055.

Kresge, Nicole, et al. “The Thermodynamic Hypothesis of Protein Folding: The Work of Christian Anfinsen.” Journal of Biological Chemistry, vol. 281, no. 14, Apr. 2006, pp. e11–13. www.jbc.org, https://doi.org/10.1016/S0021-9258(19)56522-X.

Kusters, Remy, et al. “Interdisciplinary Research in Artificial Intelligence: Challenges and Opportunities.” Frontiers in Big Data, vol. 3, Nov. 2020. Frontiers, https://doi.org/10.3389/fdata.2020.577974.

Lu, Chris, et al. The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery. arXiv:2408.06292, arXiv, 1 Sept. 2024. arXiv.org, https://doi.org/10.48550/arXiv.2408.06292.

M. Bran, Andres, et al. “Augmenting Large Language Models with Chemistry Tools.” Nature Machine Intelligence, vol. 6, no. 5, May 2024, pp. 525–35. www.nature.com, https://doi.org/10.1038/s42256-024-00832-8.

Moruzzi, Caterina. Artificial Intelligence and Creativity - Moruzzi - 2025 - Philosophy Compass - Wiley Online Library. https://compass.onlinelibrary.wiley.com/doi/10.1111/phc3.70030?af=R. Accessed 22 June 2025.

National Endowment for the Humanities. “NEH Awards $2.72 Million to Create Research Centers Examining the Cultural Implications of Artificial Intelligence.” National Endowment for the Humanities, https://www.neh.gov/news/neh-awards-272-million-create-ai-research-centers. Accessed 21 June 2025.

Nature. AI for Science 2025. www.nature.com, https://www.nature.com/articles/d42473-025-00161-3. Accessed 21 June 2025.

---. Artificial Intelligence (AI) | Nature Portfolio. https://www.nature.com/nature-portfolio/editorial-policies/ai. Accessed 21 June 2025.

NEJM AI. “NEJM AI.” NEJM AI, https://ai.nejm.org/. Accessed 21 June 2025.

Noy, Shakked, and Whitney Zhang. “Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence.” Science, vol. 381, no. 6654, July 2023, pp. 187–92. science.org (Atypon), https://doi.org/10.1126/science.adh2586.

OECD Education and Skills Today, and Skills Today. “New AI Literacy Framework to Equip Youth in an Age of AI.” OECD Education and Skills Today, 29 Apr. 2025, https://oecdedutoday.com/new-ai-literacy-framework-to-equip-youth-in-an-age-of-ai/.

O’Grady, Cathleen. Low-Quality Papers Are Surging by Exploiting Public Data Sets and AI. https://www.science.org/content/article/low-quality-papers-are-surging-exploiting-public-data-sets-and-ai. Accessed 21 June 2025.

Papagiannidis, Emmanouil, et al. “Responsible Artificial Intelligence Governance: A Review and Research Framework.” The Journal of Strategic Information Systems, vol. 34, no. 2, June 2025, p. 101885. ScienceDirect, https://doi.org/10.1016/j.jsis.2024.101885.

Perkins, Mike, et al. “The AI Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment.” Journal of University Teaching and Learning Practice, vol. 21, no. 06, Apr. 2024. arXiv.org, https://doi.org/10.53761/q3azde36.

Rep. Schiff, Adam B. [D-CA-28. Text - H.R.7913 - 118th Congress (2023-2024): Generative AI Copyright Disclosure Act of 2024. 9 Apr. 2024, https://www.congress.gov/bill/118th-congress/house-bill/7913/text. 2024-04-09.

Robert, Jenay. “2024 EDUCAUSE AI Landscape Study.” EDUCAUSE, https://www.educause.edu/ecar/research-publications/2024/2024-educause-ai-landscape-study/introduction-and-key-findings. Accessed 21 June 2025.

Rowsell, Juliette. “Flood of AI-Assisted Research ‘Weakening Quality of Science.’” Times Higher Education (THE), 12 May 2025, https://www.timeshighereducation.com/news/flood-ai-assisted-research-weakening-quality-science.

Royal Society of Chemistry. “Digital Discovery.” Royal Society of Chemistry, https://www.rsc.org/publishing/journals/digital-discovery. Accessed 21 June 2025.

Sakana AI. Sakana AI. 12 Mar. 2025, https://sakana.ai/.

---. The Al Scientist Generates Its First Peer-Reviewed Scientific Publication. 12 Mar. 2025, https://sakana.ai/.

Sam Penti, Regina, et al. Can AI Inventions Be Patented? The USPTO Speaks. | Insights | Ropes & Gray LLP. https://www.ropesgray.com/en/insights/alerts/2024/02/can-ai-inventions-be-patented-the-uspto-speaks. Accessed 21 June 2025.

Schleiger, Emma, et al. “Collaborative Intelligence: A Scoping Review Of Current Applications.” Applied Artificial Intelligence, vol. 38, no. 1, Dec. 2024, p. 2327890. Taylor and Francis+NEJM, https://doi.org/10.1080/08839514.2024.2327890.

Schmidgall, Samuel, et al. Agent Laboratory: Using LLM Agents as Research Assistants. arXiv:2501.04227, arXiv, 8 Jan. 2025. arXiv.org, https://doi.org/10.48550/arXiv.2501.04227.

Schmidt, Eric. “Opinion: AI Will Change How Universities Do Scientific Research.” GovTech, 6 July 2023, https://www.govtech.com/education/higher-ed/opinion-ai-will-change-how-universities-do-scientific-research.

Science. Science Journals: Editorial Policies. https://www.science.org/content/page/science-journals-editorial-policies. Accessed 21 June 2025.

Si, Chenglei, et al. Can LLMs Generate Novel Research Ideas? A Large-Scale Human Study with 100+ NLP Researchers. arXiv:2409.04109, arXiv, 6 Sept. 2024. arXiv.org, https://doi.org/10.48550/arXiv.2409.04109.

Spencer Foundation. Rapid Response Bridge Funding Program. https://www.spencer.org/grant_types/rapid-response-bridge-funding-program. Accessed 21 June 2025.

Stanford University. Understanding AI Literacy | Teaching Commons. https://teachingcommons.stanford.edu/teaching-guides/artificial-intelligence-teaching-guide/understanding-ai-literacy. Accessed 22 June 2025.

The AI Security Institute. “Grants | The AI Security Institute (AISI).” The AI Security Institute, https://www.aisi.gov.uk/grants. Accessed 21 June 2025.

UNESCO. Artificial Intelligence in Education | UNESCO. https://www.unesco.org/en/digital-education/artificial-intelligence. Accessed 22 June 2025.

United States Patent and Trademark Office. USPTO Issues Inventorship Guidance and Examples for AI-Assisted Inventions. https://www.uspto.gov/subscription-center/2024/uspto-issues-inventorship-guidance-and-examples-ai-assisted-inventions. Accessed 21 June 2025.

University of Toronto School of Graduate Studies. Guidance on the Appropriate Use of Generative Artificial Intelligence in Graduate Theses – School of Graduate Studies. 31

Mar. 2025, https://www.sgs.utoronto.ca/about/guidance-on-the-use-of-generative-artificial-intelligence/.

U.S. National Science Foundation. Artificial Intelligence | NSF - National Science Foundation. 17 June 2025, https://www.nsf.gov/focus-areas/artificial-intelligence.

U.S. National Science Foundation. Rapidly Accelerating Research on Artificial Intelligence in K-12 Education in Formal and Informal Settings | NSF - National Science Foundation. 8 May 2023, https://www.nsf.gov/funding/opportunities/dcl-rapidly-accelerating-research-artificial-intelligence-k-12/nsf23-097.

UT Southwestern Medical Center. LibGuides: Artificial Intelligence (AI) Guide: Science Journals. https://utsouthwestern.libguides.com/artificial-intelligence/ai-publishing-science. Accessed 21 June 2025.

Varadi, Mihaly, et al. "AlphaFold Protein Structure Database in 2024: Providing Structure Coverage for over 214 Million Protein Sequences." Nucleic Acids Research, vol. 52, no. D1, Jan. 2024, pp. D368–75. Silverchair, https://doi.org/10.1093/nar/gkad1011.

Vidal, Kathi. AI and Inventorship Guidance: Incentivizing Human Ingenuity and Investment in AI-Assisted Inventions. 12 Feb. 2024, https://www.uspto.gov/blog/ai-and-inventorship-guidance-incentivizing.

Wen, Jiaxin, et al. Predicting Empirical AI Research Outcomes with Language Models. arXiv:2506.00794, arXiv, 1 June 2025. arXiv.org, https://doi.org/10.48550/arXiv.2506.00794.

Winsberg, Eric. "Computer Simulations in Science." The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta and Uri Nodelman, Winter 2022, Metaphysics Research Lab,

Stanford University, 2022. Stanford Encyclopedia of Philosophy, https://plato.stanford.edu/archives/win2022/entries/simulations-science/.

World Health Organization. WHO Releases AI Ethics and Governance Guidance for Large Multi-Modal Models. https://www.who.int/news/item/18-01-2024-who-releases-ai-ethics-and-governance-guidance-for-large-multi-modal-models. Accessed 22 June 2025.