



Original Research Paper

Artificial Intelligence Applications in Contemporary Multidisciplinary Research with Management Development

Nayani Sateesh¹, Prof. Dr. P. Raja², Dr. Ajab Singh Choudhary³, Dr. Jahnabi Hazarika⁴, Dr. Tribeni Saikia⁵

¹Senior Assistant Professor, Computer Science and Engineering, CVR College of Engineering, Ranga Reddy, Hyderabad - 501510, Telangana, India

Email: n.sateesh@cvr.ac.in

²Associate Professor, SRM Madurai College for Engineering and Technology

Email: rajampr@gmail.com

³Assistant Professor, Department of Medical Laboratory Technology, SOAHS, Noida International University, Gautam Budh Nagar, Greater Noida

Email: ajab.singh@niu.edu.in

ORCID: <https://orcid.org/0000-0002-2338-9106>

⁴Assistant Professor, Department of Education, Mahapurusha Srimanta Sankaradeva Vishwavidyalaya, Nagaon

Email: jahnabi013@gmail.com

⁵Associate Professor, Department of Education, Mahapurusha Srimanta Sankaradeva Vishwavidyalaya, Nagaon

Email: saikia.tribeni76@gmail.com

ABSTRACT

Analysis Artificial intelligence can increase the analytical ability of multidisciplinary studies, but its role varies with the way institutions build the managerial judgment, integrate expertise, and test evidence. The literature based study looks at the applications in fields of scientific discovery, environmental analysis, social research and knowledge work in organizations. It links these applications to management development which becomes the purposeful enhancement of leadership potential, decision making and organizational learning. The choice of publications is examined in terms of a comparative framework discriminating between predictive performance, shown task benefits, and hypothetical outcomes in the institutions. The examples published support the fact that artificial intelligence can assist specialized prediction and improve some functions of knowledge; they do not insist on the improvement of the quality of research or managerial abilities in general. The main hypothesis is that technical capability can be converted to the practice of responsible research through a mediation process of management development. A proposed implementation framework, therefore, is a synthesis of disciplinary validation, shared responsibility of data, orchestrated experimentation and mindful learning. Three comparative tables are generated from this argument which represents application choices, evidence boundaries, and evaluation controls. The analysis provides a practical plan to evaluate the scientific utility as well as the further evolution of human experience. No primary survey, organizational intervention or computational experiment was carried out.

KEYWORDS: Artificial intelligence; multidisciplinary research; management development; organizational learning; research governance; human judgment.

*Corresponding Author:

Received: 25th July, 2026 | Revised: 26th August, 2026 | Accepted: 18th September, 2026 | Available Online: 24th September, 2026

(DOI): 10.70102/AEJ.2026.18.7s.38

1. Introduction

1.1. Background of the Study

Multidisciplinary research introduces different varieties of expertise to issues that are beyond the explicatory power of single disciplines. A health project in and through the community can be a combination of biological data, population rates, the interpretation of behavior and allocation of resources. The environmental project might entail physical modelling in addition to economic analysis and the knowledge of the people. Artificial intelligence is applicable in cases where these collaborations deal with large data volumes, more complicated associations, or safer analytical tasks. It is a helpful tool and it is determined by the research question and the appropriateness of the available evidence.

Wang et al. (2023) explain that applications are used in the scientific discovery process, such as the generation of hypotheses, experimental design, and analysis of heterogeneous data. They describe an application that is useful not only to conversational systems. Representation learning, predictive models and generative methods address various issues. This difference is important to management development since managers should be aware of the actual contribution of a system before entrusting power, acquiring infrastructure or measuring outcome.

In this case, management development entails individual education and institutional capacity development. Technical knowledge is needed to challenge evidence, disciplinary knowledge on the general ability to identify contested assumptions, and orchestration expertise to mobilise responsibilities. Management is also practiced by researchers when they focus on the experimental process, oversight of assistants, address data access, or set publication criteria. As a result, the managerial development is considered as a research design and not as an independent administrative activity.

1.2. Problem Statement

It is common to find that adoption judgments often entail a combination of three expectations; higher speed of work, enhanced scientific inferences as well as greater organizational ability. Such consequences demand various pieces of evidence. A system that results in a plausible report in a short time might relieve drafting time but not enhance the validity of its argument. A prediction model can also be very competent on past observations and fall flat in a new institution. In addition, short-term help could mask the front of employee development of independent competence. The issue that will be discussed here is the way that multidisciplinary research organizations can bridge the gaps between artificial intelligence applications and management development and maintain these differences.

1.3. Objectives of the Study

The research includes four goals: to juxtapose representative applications in fields of research; to differentiate between proven advantages and suggested organizational impacts; to draw a framework of

managerial skills needed to engage in responsible integration; to create an assessable framework where adoption, scientific validation, and professional development intersect.

1.4. Research Questions

What research can be assisted by artificial intelligence? Which is the evidence that an application could be transferred across disciplinary or organizational boundaries? What are the ways in which managers should form judgment to oversee such systems? Lastly, what can institutions do to know whether adoption enhances research quality and long-term human capability and short-term productivity?

1.5. Applications Across Research Disciplines

Protein structure prediction is a good illustration of a limited scientific usage. Jumper et al. (2021) have shown considerable advances with the help of AlphaFold. This success has to do with a specified prediction issue that has well-defined assessment protocols. It alone does not prove therapeutic effectiveness and omit the necessity to explore biological functioning. To research managers, the example demonstrates that appropriateness of investment decision is based on connecting it with particular bottlenecks of scientific decisions and relevant validation resources.

Another application is in materials research. Inorganic crystal structures of candidate structures: Merchant et al. (2023) described a deep learning experience in the exploration of candidate structures inorganic structure. Predicted stability based on computations must not have confused with effective synthesis and usable material performance. To organize such work, an institution should match the computational scientists with experts in the field of experiments, and keep a register that would link every applicant with the further assessment. The managerial issue is related to the entire research process and unsuccessful candidates and resource limitations.

The other fields need various interpretations. In social research, the language models can help to structure documents, propose coding typologies, or create tentative summaries. They can be used in design of feedback activities in education research. These are candidate applications that are analytically discussed, as opposed to predetermined enhancements in all contexts. Researchers should examine the question of whether machine generated categories can maintain the meaning of participants and whether assistance can result in greater learning than doing the task at hand. Application options and the managerial requirements are described in Table 1.

Table 1. Representative applications and proposed management development priorities.

Research domain	Candidate application	Validation requirement	Managerial capability
-----------------	-----------------------	------------------------	-----------------------

Life sciences	Protein structure prediction	Independent biological interpretation	Coordinate computational and experimental expertise
Materials research	Candidate structure screening	Synthesis and property testing	Allocate resources across uncertain candidates
Social research	Assisted document coding	Contextual review by researchers	Resolve disagreements about meaning
Education research	Feedback design assistance	Measure learning and transfer	Develop assessment literacy
Organizational research	Knowledge task assistance	Compare quality and completion time	Redesign work with employees

1.6. Analytical Capabilities and Evidence Interpretation

Prediction, generation and explanation should be assessed separately. Prediction is what you do on observations not used to build your model. When we do generation, we will come up with candidate text, designs, or hypotheses. Explanation involves reasoning about how or why something is the case or about how or why a conclusion follows. Success in one category does not mean success in another. After predicting an explanation may be inconsistent with the process from which it came. It is this that managers should therefore aim to be statistically literate. They should consider if the evaluation data were independent, if the comparability of the method was convincing and if the reported performance was expected from the group where the evaluation was conducted. They also need to understand what is meant by "a model's confidence" vs. justified scientific confidence. These capabilities enable intelligent co-operation with experts without the need for every manager to be a model developer.

1.7. Management Development and Organizational Learning

Noy and Zhang (2023) concluded in an experiment that generative artificial intelligence offers benefits in terms of productivity for professional writing tasks. Brynjolfsson et al., (2025) looked at an assistant that was implemented in customer support workers and found varying returns to different levels of experience and skills. These studies provide valuable information with respect to specific workplaces. They never directly check whether or not research managers are able to be more

leaders or whether or not multidisciplinary teams of researchers create more valid finds.

Here proposed managerial implication is to make development all about real decisions. Participants are able to make comparisons between an assisted analysis and an analysis they have prepared independently; and explain disagreements and documentation on why decisions were made to accept/reject suggestions. Supervisors are then able to assess the reasoning quality of its use instead of simply counting its use. This allows for practice in verification, expression of uncertainty and detection of when convenience is substituting for analysis.

1.8. Governance and Responsible Research Practice

The NIST Artificial Intelligence Risk Management Framework is a voluntary framework which consists of four stages—governing, mapping, measuring, and managing risk (NIST, 2023). These functions in research organizations can help to identify the groups that are affected, to determine how to evaluate and respond to problems observed, and to assign responsibility. The framework does not provide evidence of the effectiveness of a specific implementation, but rather of a governance resource.

There should be an institutional policy that defines appropriate data inputs, outlines review responsibilities and outlines what to do in the event of failure. Transactional acts concerning information from confidential interviews, unpublished research and confidential organizational documents. The proposed policy should also outline how assistance would be acknowledged in "research products". Consequential uses should be disclosed in such a way that the reader is able to see the impact they have on analysis and interpretation.

1.9. Integration and Remaining Knowledge Gaps

Artificial intelligence and its potential fostering illusions of understanding and limiting scientific inquiry is considered by Messeri and Crockett (2024). They draw attention to a danger ignored by other measures of productivity: that an organization can become more productive while undermining key relationships. In this research, therefore, intellectual diversity and independent thinking are considered as priorities for evaluation.

There is also a large gap that has to do with assistance and durable management capability. Based on past examples, little reason can be advanced for forecasting future leader performance once the helping hand will be removed, whether tensions between subjects will be resolved or whether smaller schools can continue. Together these questions are fundamental in being questions that must be answered through organizational research and not by extrapolating on technical benchmarks.

2. Results and discussion

2.1. Application Comparability and Research Value

The end product of this synthesis is a scheme for understanding evidence across application. The chosen examples vary with respect to the units of analysis, the ways of evaluating, and the purposes for which they are evaluated. Performance metric comparisons, between protein prediction and material screening and between professional writing, are quickly becoming irrelevant if done without important context. Comparability should start with the action that each application is supporting and what can be problematic if it is incorrect.

To have some research value, three conditions are proposed to be necessary. The first criterion is that there be a problem in research that necessitates the use of the application. Second, there needs to be validation of its outputs applicable to the discipline. Third, the group needs to have ability to take action based on reliable outputs and disregard unreliable outputs. These conditions are not supposed to be causal relationships estimated below in the paper, but are analytical propositions stated here.

Suppose there is a possible university partnership to assess urban heat interventions. A system using Artificial Intelligence could be used to classify documents and organize environmental observations. There would still be issues of organizing the geographic range and quality of measurement, and what it means to be vulnerable. Limitations for the implementation would have to be explained by the municipal partners in general. The example illustrates that there are other, non-technical issues that disagreements over research priorities or benefits distribution cannot be resolved by technical accuracy.

A manager in this context should keep a record of decisions associated with each application which includes who is to keep an eye on the outputs and where the application is being used and its limitations. If a model is not considered appropriate then a modifiable alternative should be retained. This means that use depends on observed usefulness and not on how enthusiastic organizations are about using them or on how they have already invested in them.

2.2. Interpretation of Performance and Human Capability

There is a need to investigate model, task level, team level and institutional level performance together. Model Evaluation: Determine if model predictions/created products satisfy given criteria. Task evaluation will take into account quality, time and the effort required to make the correct answer. When evaluating team, issues of coordination and whether responsibilities are still clear are looked at. Institutional aspect is judged, including the need for continual skills and costs for evaluations. At one place the improvement may go hand-in-hand with deterioration elsewhere.

For instance, the preliminary summaries may be produced quickly, while more time is required to ensure the accuracy of citing and solve ambiguous statements. The outcome for this is the overall effort needed to create a reliable synthesis. Furthermore, if the necessary verification is not yet settled up and paid for, a lack of drafting time will not help. The management should thus be careful not to reward such speedy writing in order to discourage careful reading.

The measurement of human capability must be done separately. An assisted task will test a performance with support, and then a follow up independent task will test for transfer. Both can be used in a development program, and they are about the same difficulty level, and are easily reproducible and easily scoreable. Depleted favourable self assessments are not a substitute for demonstrated competence, although they will aid in the explanation of choices made. These schemes might provide an empirical evaluation of the difference between assistance and development.

2.3. Evaluation of Management Development

2.3.1. Decision Quality and Professional Judgment

A proposed management development curriculum ought to start with decisions confronted by participants. These can include an interpretation of a project forecast, an appraisal of a summary of evidence, or a judgement come to by using an automated recommendation, which requires further investigation. Participants should consider what they would like to know before making a recommendation as well as situations in which a recommendation should be turned down. Assessment should reward pointing to justified reasoning – appropriate uncertainty (wagging).

There should be training based on contexts in which the system is beneficial and contexts in which it is not. Successful assistance may lead to a sense of habit for each exercise if it is given at the end of each exercise. On the other hand, activities that are all about failing may lead to false rejection. A balanced curriculum should be based on conditions of trust and evidence and suitability of task. The instructor should determine if participants are able to continuously explain conditional judgment.

2.3.2. Linking Learning to Organizational Outcomes

Development outcomes are organizationally valuable if they lead to an environment where judgment is improved and thus influences the nature of the research. These may manifest themselves in the form of better thought-out analytical plans, earlier identification of unsupported claims, fewer residual handoff errors, etc., and more complete documentation. Prior to the implementation these indicators should be defined. They need to factor in staffing, project complexity and various other things going on in the organisation that can impact on performance alone.

Additional evidence is needed to establish the development outcomes in management, whether these are published precedents or additional evidence. It is this

difference that makes it difficult to report studies of performance or productivity of technical groups as direct evidence of leadership development.

Table 2. Evidence boundaries and proposed organizational tests.

Published precedent	Supported interpretation	Additional evidence required
AlphaFold study	Progress on protein structure prediction	Effects on research coordination and judgment
Materials screening study	Computational exploration of candidate structures	Experimental usefulness and resource efficiency
Professional writing experiment	Benefits within assigned writing tasks	Durable learning and independent assessment
Customer support study	Context dependent workplace assistance benefits	Transfer to research management responsibilities

2.4. Integrated Implementation and Institutional Development

The proposed implementation process starts with a specific research need and an explicit decision as to what a specific form of help is to achieve. The assisted workflow should be compared to ongoing practice by a small pilot that uses pre-defined quality criterion. The institution could then look at whether it is scientifically useful and whether staff learning is involved and whether the overheads would be too heavy. Expansion should be based upon this joint evaluation.

XO Responsibilities should always be visible throughout. Assessment of scientific meaning is the responsibility of a disciplinary lead, assessment of model behaviour and reproducibility is the responsibility of a technical lead, and coordination of resources and decisions is the responsibility of a project manager. There may be less people in a small team that perform these roles, but the roles should still be kept separate. If they have implementation issues, these individuals should also have the opportunity to report the issue and oppose misuse.

Data preparation, evaluation, training, and maintenance should be a part of resource planning. Subscription/computing expenditure isn't the total cost of adoption. It's important to factor in review time, as well as optional workflows, for the manager. Narrow application, and well defined evaluation criteria, may be more defensible under limited resources than even broad or institution-wide implementation with ambiguous accountability measures.

Equity is dealing with the work of some that is seen and others that are not; it's thinking about people who were not appreciated for skills. Sometimes an automated summary can give over-representation to documents in their languages or following their formats. The interpretation of a missing perspective should be sought out by disciplinary/whole team/community team leaders. This is suggested as a measure of research quality and a principle of participation as excluded evidence can lead to flawed research questions.

The plan additionally indicates periodic review once you've implemented it. The usefulness can change over time due to changes in data, models, staff experiences, or research goals. It is important to review together what works and what doesn't. Any changes or ending of a system should be documented with reasons for such change. Helps retain institutional memory and provides future teams evidence for why previous decisions were made.

Another differentiation has to do with the multidisciplinary involvement and the real integration. Several Departments can work on one project, and all decisions can be made by the assumptions made by one discipline. In the proposed framework the teams have to define the points of difference between their ideas before merging their output. For example, a description that makes sense in a statistical analysis might not be sufficient as an interpretive researcher with interests in context. Managers should be explicit about such disagreements and designate who has authority to approve changing of common definitions. These negotiations provide useful training exercises for development activities in translating knowledge. The goal is to help the team reach less law-defying or more law-friendly resolutions of disagreements that do not resort to all knowledge into knowledge that fits a certain analytical system.

Getting procurement to do its job can also be a learning experience for management. Managers should clarify the proposed task, the findings needed to make a judgement about success and the situations in which use should be discontinued prior to choosing a system. Teams can correlate a specialized application with easier analytical approaches as well as human workflows that are already in place. Such comparison should also take into account compatibility with the research records, access arrangements as well as meaningful evaluation data. While technically impressive, a demonstration where the tasks the person has selected are different from their local tasks, provides little guidance for local use. When the selection rationale is recorded, review can take place later and it becomes easier to differentiate between an unsuitable application and an application where the preparation and/or support has been inadequate.

The approach that is being proposed is that disagreement between humans and systems is information. If reviewers are continually rejecting an output, review the work to see if it's a problem with content inaccuracy, objective,

unfinished data, or mismatch of disciplinary practice. There are different answers that are suggested by these explanations. If the system is answering the wrong question of a user, then it may need to perform the task differently, or simply, if it is acting as though it didn't understand the user, it means that training is required. One such development activity is for participants to diagnose the disagreement first, before suggesting a remedy. This will help managers to examine the processes that have failed, and not just blame workers or technology.

An additional suggestion for research is how learning opportunities are distributed. In the case of delegation, where experienced researchers might undertake all analysis tasks preliminary to them, younger researchers might lose out on the opportunities they normally had to learn to recognise patterns and limitations. Otherwise, support could enable supervisors to have more time to have in-depth discussions of challenging situations. Teams should report what activities can continue to be done for practice and if there are changes with mentoring post-adoption. The process of development planning can be intentional and may involve a decision, selection, and continuation of specific independent tasks with the support of assistive development for other tasks. An evaluation can then be done to determine if this structure facilitates further progress to more rigorous duties, and continued disciplinary effectiveness.

Lastly, institutional assessment needs to differentiate between good program ending and organisms losing interest. Knowledge can still be gained from a pilot, even when the benefits are limited and/or review costs large. Managers should conclude that there are lessons learned, there were things that were discovered to be assumptions that were incorrect, and if the application scope was narrowed should it still be tested? Only funding arrangements that reward expansion can make it difficult to report these conclusions. The proposed framework instead suggests that decisions to discontinue should be considered as appropriate research management decisions where there is documentation of why that decision was made. This makes it easier to report in an honest way – a team cannot be tempted to spin challenges as successes or spin less than clear results as successful as they can when the numbers are not derived from their own performance. It is also a resource conservation for projects where future benefits continue to be evident.

Additionally, there must be a clear and fair process for raising queries when decisions are made based on automated results. A researcher needs to understand who to reach out to if they think that there is a discrepancy between a recommendation and what they have seen/and their professional duties. The decision should be recorded on the part of the responsible manager, the circumstances must be reviewed as necessary and the result must be conveyed. Of particular interest is studying such transfers or exchanges and how they might strengthen the collective reasoning or simply shift the decision-making

authority to a system whose recommendations team members may find it hard to challenge in practice.

3. Conclusion

3.1. Conclusions

When opportunities are accompanied by a limited capacity task, and assessed by the specified evidence, AI can provide helpful tools for multi-disciplinary research. Specific scientific support was found for the publications reviewed, with a less sure general claim for management development. Accordingly, the paper proposes that managerial judgment, co-ordinated responsibility and organizational learning be also accorded a due consideration.

The contribution is a connected framework which establishes a link between application choices, disciplinary legitimation, and professional development. Research institutions need to consider whether support equates to reliable performance, greater independence and improved opportunity for making responsible choices. Through this, adoption is a process that is ongoing and can be expanded, revised or withdrawn due to specific reasons.

3.2. Limitations of the Study

In this study a narrative synthesis approach was used, and it was selective. It should not have sweeping insights, measure a typical impact, or suggest cause-and-effect relationships between the use of AI and management development. There is a lack of commonality in the technologies and settings of the selected studies. Institutions recommendations are simply "suggestions" based on reason and not necessarily tested empirically. The relevance of specific uses may also change very quickly with changes in technology, however, the concept of the task performed and a more permanent capability is useful for analytical purposes.

3.3. Future Scope of Research

In the future, it is recommended to conduct a comparison study of two different management development programs with and without the support of artificial intelligence in various institutions. Independent judgement should be assessed prior to training, whilst working independently, and when training support is terminated. A longer observation period may be evaluated for documentation, collaboration and the quality of research efficacy and dependence. Work needs to be done and compared in resource-constrained institutions and multi-lingual teams. Qualitative interviews would be useful to give greater insight to the variation in organizational outcomes from similar technologies, and transparency of the reporting would help be replicated and fine-tuned the proposed framework.

4. Material and methods

4.1. Source Materials and Evidence Selection

A purposive selection of scholarly publications and institutional governance were implemented. Sources were selected based on if they met one of the following criteria: Scientific applications; workplace assistance; or

interpretation of research risks. Publisher, institutional or author accessible records of bibliographic information and main claims were verified. The evidence set is comprised of both empirical investigations and conceptual discussion; the roles of each of these are distinguished in the analysis.

In contrast, this selection strategy facilitates the comparison of contrasting use of applications but does not ensure the completeness of the selection. Systematic review process screening was not performed, nor database search and registered review process, nor a comprehensive eligibility assessment. Sources have been cited to support specific claims, not to suggest a field of authority. There are no newly-collected participant information or measurements from the deployments of the organizations found in the manuscript.

4.2. Comparative Analytical Design

Every source was relevantly analyzed using the matrix with the following dimensions: application, research setting, unit of analysis, contribution reported and limits of inference. The synthesis subsequently matched those functions with the abilities a manager would want to manage a comparable research workflow. The proposed curriculum, governance and evaluation model for the paper was removed from the published results. This ensured that the recommendations were not presented as actual results.

The organization of cross disciplinary comparison emphasized the commonality of questions instead of commonality of the numerical effects. These questions involved the suitability of the data, validation questions and questions about consequences of error and allocation of responsibility. Scientific usefulness and organisational convenience were kept as subject matter for investigation where there was any conflict between them. No effect sizes were pooled due to studies selected not having a similar outcome or intervention.

4.3. Proposed Implementation Procedure

In a future institutional study, the teams involved, the tasks to be carried out and the applications allowed had to be first of all defined. A baseline test should be conducted before the commencement of the training. Groups need to be provided with clear instructions to check outputs and research information was protected. The evaluation should leave sufficient detail in terms of version of the system, instructions placed on the tasks, and intervention with humans to account for variations between participating groups.

Teams should be matched to similar conditions of the assists, and conditions being compared to each other, where possible. If random assignment cannot be accomplished, then the investigators will need to explain the comparison method and include differences that could impact the comparison results. Implementation records should be able to differentiate between access to a system and actual use of a system. They ought to likewise take down interruptions, task adjustments, and deviations from the organised activity.

4.4. Outcome Assessment and Analytical Controls

Primary outcomes should be work quality which is independently rated, and total completion time (including verification and correction). Independent judgment, quality of collaboration and thoroughness of documentation may represent secondary outcomes. Assignments should be judged by clear standards and judges should not know the assigned condition of the studies if possible. The project should gauge if there is any such variation, between experience, discipline and level of complexity in the nature of the tasks, without necessarily assuming that the patterns of exploratory subgroups constitute findings.

Analysis should separate successive observations from different teams of participants/trial of participant(s). Plan for an appropriate sample size that is dependent on the primary outcome and study design. Non-exposure should be reported as missing observations, withdrawals or changes in system access. Qualitative accounts may be used to describe the implementation experiences, but should not be indistinguishable from quantitative effect estimates. Table 3 provides Proposed Controls for this future evaluation.

Table 3. Proposed evaluation controls and their interpretive purposes.

Evaluation control	Purpose	Required record
Baseline comparison	Identify starting differences	Initial quality and experience
Independent rating	Reduce assessment bias	Scoring criteria and reviewer agreement
Complete time accounting	Capture hidden correction effort	Drafting and verification duration
Unaided follow up	Assess retained competence	Independent task performance
Version documentation	Explain technological changes	Model and instruction records
Failure reporting	Prevent selective interpretation	Errors and corrective actions

References

1. Brynjolfsson, E., Li, D., and Raymond, L. R. (2025). Generative AI at work. *The Quarterly Journal of Economics*, 140, 889–942. <https://doi.org/10.1093/qje/qjae044>
2. Jumper, J., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
3. Merchant, A., et al. (2023). Scaling deep learning for materials discovery. *Nature*, 624, 80–85. <https://doi.org/10.1038/s41586-023-06735-9>
4. Messeri, L., and Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. *Nature*, 627, 49–58. <https://doi.org/10.1038/s41586-024-07146-0>
5. National Institute of Standards and Technology (NIST). (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100–1. <https://doi.org/10.6028/NIST.AI.100-1>
6. Noy, S., and Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381, 187–192. <https://doi.org/10.1126/science.adh2586>
7. Wang, H., et al. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, 620, 47–60. <https://doi.org/10.1038/s41586-023-06221-2>