American Journal of Advanced Multidisciplinary Research and Innovation (AJAMRI)

E-ISSN: XXXX-XXXX ● Website: www.ajamri.com ● Email: editor@ajamri.com

AI and Human Expertise in Scientific Decision-Making: Designing Effective Human–Machine Research Partnerships

Dr. Sophia Eleanor Hartmann

Associate Professor, Technical University of Munich, Germany

Abstract

Artificial intelligence is becoming an important participant in scientific decision-making. Machine-learning systems can search extensive bodies of literature, detect complex patterns, identify promising hypotheses, prioritize experiments, estimate uncertainty, and assist with the interpretation of multidimensional observations. These capabilities can expand the analytical capacity of research teams, yet they do not independently provide the contextual understanding, methodological judgment, causal reasoning, ethical awareness, or accountability required for responsible scientific inquiry. The central challenge is therefore not whether artificial intelligence should replace scientific expertise, but how researchers and intelligent systems can be organized into effective decision partnerships.

This study develops a simulation-based framework for comparing five scientific decision arrangements: human-only analysis, AI-only analysis, AI recommendations followed by human review, explainable human–AI collaboration, and adaptive human–AI partnership. Each arrangement was assessed through a composite decision-quality index incorporating analytical accuracy, contextual validity, interpretability, error detection, uncertainty management, reproducibility, and accountability. The resulting simulated scores were 71, 76, 84, 89, and 92, respectively. These values are illustrative scenario outputs rather than observed experimental results. The comparison indicates that neither unaided human judgment nor independent machine analysis achieved the strongest overall performance. Higher scores emerged when computational scale was combined with expert interpretation, explanation-based review, structured disagreement, and dynamically allocated decision authority.

The study concludes that effective scientific partnerships require complementary rather than merely additive interaction. Artificial intelligence should be assigned tasks suited to rapid computation, pattern recognition, and evidence retrieval, while researchers should retain responsibility for problem formulation, construct validity, causal interpretation, ethical judgment, and final scientific claims. Reliable partnership also depends on transparent documentation, calibrated trust, data provenance, reproducible workflows, and mechanisms for contesting machine recommendations. Properly designed human–machine research partnerships can improve decision quality while preserving the epistemic responsibility at the center of scientific practice.

American Journal of Advanced Multidisciplinary Research and Innovation (AJAMRI)

E-ISSN: XXXX-XXXX ● Website: www.ajamri.com ● Email: editor@ajamri.com

***Keywords*:** Human–AI collaboration; scientific decision-making; research partnership; augmented intelligence; explainable artificial intelligence; scientific expertise; research integrity; machine-assisted discovery

1. Introduction

Scientific inquiry involves a continuous sequence of decisions. Researchers select problems, define constructs, evaluate theories, design experiments, determine sampling strategies, inspect data quality, choose analytical methods, interpret uncertain findings, and decide whether the evidence supports a defensible conclusion. These decisions are not purely computational. They are shaped by disciplinary knowledge, methodological norms, tacit experience, ethical considerations, material constraints, and awareness of the historical development of a research field. Even where a statistical or computational procedure can be automated, the scientific meaning of its output continues to depend on human interpretation.

Artificial intelligence has expanded the scale at which many research activities can be performed. Machine-learning systems can process datasets that exceed unaided human capacity, identify nonlinear relationships, compare large numbers of candidate models, automate image or signal analysis, and retrieve potentially relevant literature. In chemistry and materials science, algorithms can prioritize candidate compounds or structures. In biomedicine, intelligent systems can analyze molecular, imaging, clinical, and genomic information. In environmental science, they can integrate satellite observations, sensor streams, simulations, and historical records. These capabilities can shorten parts of the research cycle and make previously impractical analyses achievable.

The expansion of computational capability does not eliminate scientific uncertainty. An algorithm can learn a strong association from a dataset while relying on a confounding feature, an institutional artifact, or a measurement error. Schramowski and colleagues demonstrated the importance of allowing scientists to interact with model explanations so that models can be corrected when they learn for the wrong scientific reasons. Similarly, Jiménez-Luna, Grisoni, and Schneider emphasized that explainability is particularly important in drug discovery because predictive performance alone may not provide a meaningful scientific account of molecular behavior.

Human reasoning also has limitations. Researchers are susceptible to anchoring, confirmation bias, selective attention, disciplinary conventions, and inconsistent judgment. Large evidence spaces may exceed an individual's memory and processing capacity. Decisions made under time pressure may rely excessively on familiar theories or methods. Artificial intelligence can help expose overlooked evidence, test alternative specifications, and identify inconsistencies. However, it can also amplify biases embedded in training data, generate unjustified confidence, obscure provenance, or encourage automation bias among users.

The appropriate scientific question is therefore not whether humans or machines are inherently superior. Their relative value changes across tasks and conditions. Artificial intelligence may outperform researchers in repetitive pattern recognition but remain weak at assessing whether a variable meaningfully represents a theoretical construct. Experts may recognize contextual anomalies but overlook weak statistical relationships distributed across millions of observations. This study

investigates how their capabilities can be coordinated within a research partnership that improves decision quality without weakening scientific accountability.

2. Review of Literature

2.1 Complementary Intelligence in Decision-Making

Research on human–AI collaboration frequently emphasizes complementarity. Jarrahi described human–AI symbiosis as a decision arrangement in which computational systems contribute analytical power while humans provide intuition, contextual understanding, and holistic judgment. This position differs from complete automation because it treats artificial intelligence as one participant within a broader decision structure. The value of the partnership depends on whether tasks are allocated according to comparative capability.

Human expertise develops through formal training, repeated exposure, feedback, and participation in disciplinary communities. Experts often recognize patterns without being able to describe every cognitive step involved. This tacit knowledge can be especially important when information is incomplete or the situation differs from previously documented cases. Yet expertise can become rigid when familiar patterns dominate interpretation. Machine-generated alternatives may help experts reconsider assumptions, provided that recommendations are presented in a form that permits meaningful evaluation.

Artificial intelligence contributes consistency, computational speed, and the ability to analyze large evidence spaces. It can compare thousands of candidate explanations or identify subtle correlations across heterogeneous datasets. Nevertheless, its competence is bounded by objectives, representations, training data, and evaluation procedures. A model can optimize the wrong target with great precision. Complementarity therefore emerges only when the human partner can recognize whether the machine's objective and evidence are scientifically appropriate.

2.2 Explainability, Trust, and Scientific Understanding

Trust is central to human–machine collaboration, but high trust is not always desirable. Researchers should neither reject artificial intelligence merely because its reasoning is unfamiliar nor accept it simply because its output appears statistically sophisticated. Appropriate reliance requires calibrated trust: confidence that increases when the system performs reliably within its documented boundaries and decreases when evidence becomes uncertain, anomalous, or outside the training distribution.

Amershi and colleagues developed guidelines for human–AI interaction that emphasize setting expectations, making clear what a system can do, supporting efficient correction, and communicating changes in behavior. These principles have direct scientific relevance. Researchers need to know whether a model is producing a prediction, ranking, similarity measure, generated hypothesis, or causal estimate. They also need to understand the conditions under which the output becomes unreliable.

Explainability can facilitate scientific evaluation by showing influential variables, comparable cases, uncertainty ranges, or alternative predictions. Nevertheless, an explanation can be technically plausible without being scientifically faithful. Feature-importance visualizations may be unstable, and post hoc explanations may not accurately represent the model's internal process. Rudin argued that in high-stakes settings, inherently interpretable models should sometimes be preferred to explanations of

opaque systems. Scientific partnerships should consequently match the level of interpretability to the consequences of error and the purpose of the analysis.

2.3 Human–AI Collaboration in the Research Cycle

Artificial intelligence can support literature discovery by identifying related studies, clustering themes, extracting concepts, and connecting previously separate research domains. It can assist hypothesis development by identifying anomalies or predicting relationships that deserve experimental attention. During study design, it can support power analysis, simulation, optimization, and instrument calibration. During analysis, it can process images, language, sensor data, and high-dimensional measurements.

In drug discovery, explainable artificial intelligence has been used to connect molecular predictions with chemically meaningful features. In plant phenotyping, explanatory interactive learning has allowed domain experts to correct models that relied on confounding visual information. These applications illustrate an important distinction: scientists should not merely inspect final outputs. They can provide feedback during model development and thereby influence what the model learns.

The literature also identifies risks. Automated literature systems may reproduce citation bias or overlook non-digitized evidence. Generative models may produce plausible but unsupported scientific statements. Prediction models may become unreliable when applied beyond their validated domain. Collaborative systems may encourage researchers to accept machine suggestions without sufficient independent examination. The design of partnership must therefore include error detection, disagreement procedures, documentation, and human authority over consequential conclusions.

3. Objectives of the Study

3.1 Scientific and Analytical Objectives

The primary objective is to compare alternative human–AI decision arrangements and assess how progressively structured collaboration may influence scientific decision quality. The study considers whether a system that combines computational analysis with expert review, explanation, and adaptive task allocation can provide stronger performance than human-only or AI-only processes.

A second objective is to identify the distinctive contributions of human expertise and artificial intelligence. This includes examining which research tasks are suited to machine processing, which require expert judgment, and which benefit most from iterative interaction. The analysis treats complementarity as a designed property rather than an automatic consequence of placing a researcher and an algorithm in the same workflow.

3.2 Partnership-Design Objectives

The study seeks to develop principles for calibrated trust, transparent explanation, productive disagreement, and accountable decision authority. A well-designed partnership should allow researchers to challenge machine recommendations and should permit the system to reveal evidence that conflicts with established expert expectations.

The final objective is to connect technical collaboration with research integrity. Human–AI workflows must support provenance, reproducibility, authorship transparency, methodological

disclosure, and verification of generated material. The study therefore evaluates scientific decision quality through both analytical performance and responsible research practice.

4. Research Methodology

4.1 Research Design

A simulation-based comparative design was employed. No human participants, laboratories, institutional research teams, or proprietary artificial intelligence systems were involved. The numerical values were constructed to illustrate how different partnership models could be evaluated through a multidimensional decision framework. They must not be presented as observed performance data.

Five decision arrangements were considered: human-only analysis, AI-only analysis, AI recommendations reviewed by a scientist, explainable human–AI partnership, and adaptive human–AI partnership. The arrangements represent increasing degrees of interaction rather than successive versions of a single real system.

4.2 Partnership Configurations

The human-only configuration relied on expert judgment, conventional analytical tools, and disciplinary knowledge without machine-generated recommendations. The AI-only configuration represented automated pattern detection and recommendation without substantive expert review. The recommendation-with-review configuration allowed a scientist to approve, reject, or modify machine output.

The explainable partnership added evidence summaries, uncertainty indicators, feature contributions, and structured opportunities for correction. The adaptive partnership dynamically allocated tasks according to confidence, novelty, consequence, and available expertise. It also included independent human verification for high-consequence decisions and machine-assisted challenge of potentially biased human judgments.

Table 1. Functional design of the simulated research partnerships

Partnership model	Machine contribution	Human contribution	Principal vulnerability
Human only	Conventional computational support	Full formulation, analysis, and interpretation	Cognitive bias and limited analytical scale
AI only	Automated search, modeling, and recommendation	Minimal intervention	Weak contextual and ethical judgment
AI with human review	Evidence processing and ranked recommendations	Acceptance, correction, and final decision	Automation bias or superficial review

American Journal of Advanced Multidisciplinary Research and Innovation (AJAMRI)

E-ISSN: XXXX-XXXX ● Website: www.ajamri.com ● Email: editor@ajamri.com

Partnership model	Machine contribution	Human contribution	Principal vulnerability
Explainable partnership	Recommendations, explanations, and uncertainty	Interpretive evaluation and model correction	Misleading or unstable explanations
Adaptive partnership	Dynamic analysis and task allocation	Contextual judgment, escalation, and accountability	Greater coordination and governance burden

Note: These configurations are conceptual and do not describe a specific commercial system.

4.3 Measurement Framework

Scientific decision quality was evaluated through seven dimensions: analytical accuracy, contextual validity, interpretability, error detection, uncertainty management, reproducibility, and accountability. Each dimension was assigned a five-point score.

The composite scientific decision-quality index was calculated as:

$$SDQIj=100(0.20Aj+0.17Cj+0.13Ij+0.15Ej+0.13Uj+0.12Rj+0.10Gj)/5$$

where SDQIj is the index for partnership model j; A represents analytical accuracy; C, contextual validity; I, interpretability; E, error detection; U, uncertainty management; R, reproducibility; and G, accountable governance.

Analytical accuracy and contextual validity received the largest weights because a scientifically useful decision must be both technically credible and meaningful within its domain. Error detection received a substantial weight because collaborative advantage is most valuable when one partner can identify the other's mistakes.

4.4 Scenario and Analytical Procedure

The simulated setting involved an interdisciplinary research team assessing competing explanations for an unexpected pattern in a heterogeneous dataset. The artificial intelligence component could search evidence, estimate relationships, compare models, and identify anomalies. The human component could examine theoretical plausibility, measurement validity, ethical implications, and the possibility of confounding.

The analysis proceeded through configuration definition, criterion scoring, weighted aggregation, comparison, and sensitivity interpretation. Inferential tests were not applied because the values were constructed rather than sampled. A real-world validation would require preregistered hypotheses, verified datasets, repeated decision tasks, independent expert panels, baseline controls, confidence intervals, and documented model versions.

American Journal of Advanced Multidisciplinary Research and Innovation (AJAMRI)
E-ISSN: XXXX-XXXX ● Website: www.ajamri.com ● Email: editor@ajamri.com

5. Analysis

5.1 Comparative Decision-Quality Results

The human-only configuration achieved a simulated index of 71. It performed strongly on contextual interpretation, ethical awareness, and accountability but was constrained by limited analytical scale, inconsistency, and vulnerability to cognitive bias. This result does not suggest that human expertise is generally weak. It demonstrates that complex scientific evidence can exceed the processing capacity of unaided individuals or small teams.

The AI-only arrangement achieved 76. It benefited from consistent computation, rapid comparison, and broad evidence processing. Its limitations emerged in construct validity, contextual judgment, interpretability, and responsibility for the final conclusion. An algorithm could identify a statistical pattern without determining whether the pattern reflected a scientifically meaningful mechanism.

Human review increased the simulated index to 84 by allowing machine-generated recommendations to be evaluated against disciplinary knowledge. Explainable partnership reached 89 because the researcher could examine supporting evidence, uncertainty, and potential model dependencies. The adaptive partnership obtained the highest score of 92 by assigning authority according to task characteristics and requiring escalation when confidence was low or consequences were substantial.

Table 2. Simulated scientific decision-quality assessment

Partnership model	Analytical capacity	Contextual validity	Error detection	Composite index
Human only	3.5/5	4.3/5	3.2/5	71
AI only	4.5/5	2.8/5	3.4/5	76
AI recommendation with human review	4.5/5	4.1/5	4.0/5	84
Explainable human–AI partnership	4.6/5	4.5/5	4.5/5	89
Adaptive human–AI partnership	4.7/5	4.7/5	4.7/5	92

Source: Author-developed simulation based on literature-informed relationships.
Caution: Scores are illustrative scenario values and are not empirical benchmark results.

5.2 Complementarity and Error Correction

The largest simulated advantage emerged when the partnership supported mutual error correction. The machine could challenge the researcher by revealing overlooked evidence, anomalous observations, or

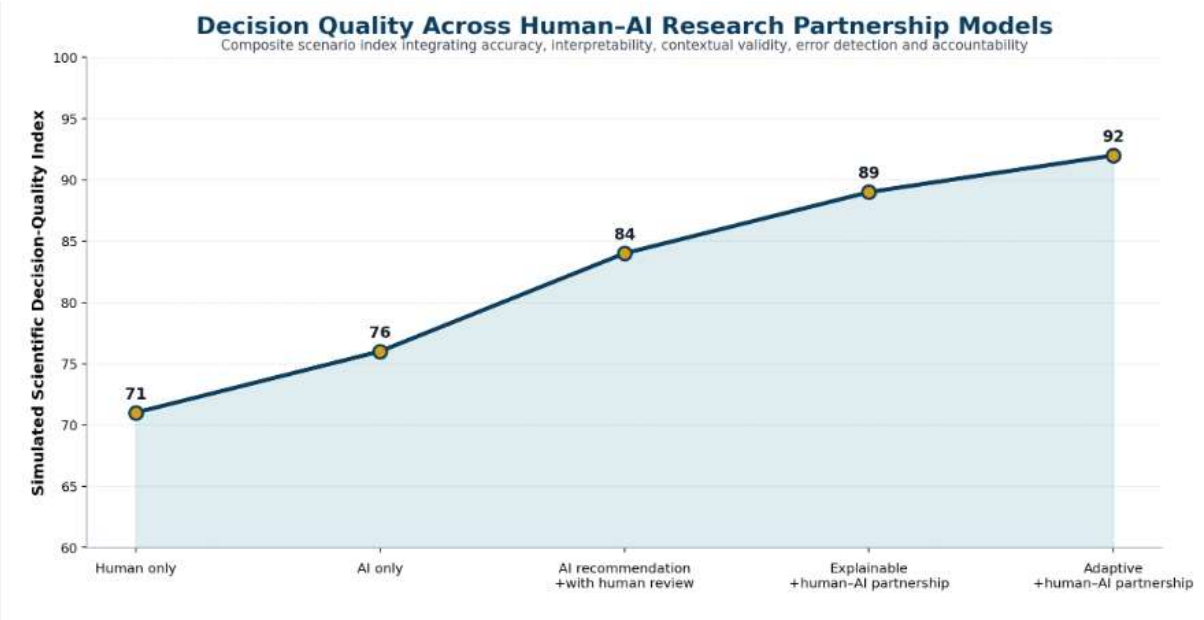
alternative models. The scientist could challenge the system by identifying confounding features, implausible mechanisms, invalid measurements, or ethically unacceptable recommendations.

This mutual correction differs from a workflow in which the machine produces a recommendation and the researcher merely approves it. Superficial review provides limited protection against automation bias. Effective review requires access to evidence, uncertainty, model limitations, and reasonable alternatives. The system should also preserve the researcher's initial judgment where possible so that independent reasoning is not immediately anchored by the machine recommendation.

The adaptive partnership performed most strongly because it did not assign permanent superiority to either participant. Routine high-volume screening was allocated primarily to the artificial intelligence system. Novel, ambiguous, ethically consequential, or theory-dependent decisions were escalated to researchers. Disagreement triggered additional evidence collection rather than automatic acceptance of the statistically more confident participant.

5.3 Graphical Presentation

Graph 1. Decision Quality Across Human–AI Research Partnership Models



Supporting data: Human-only decision-making, 71; AI-only decision-making, 76; AI recommendation with human review, 84; explainable human–AI partnership, 89; and adaptive human–AI partnership, 92

Interpretation: The graph indicates that machine analysis alone produced a modest simulated improvement over unaided human analysis. The larger gain occurred when computational recommendations were combined with human review. Explanation and adaptive allocation produced further improvements by strengthening contextual evaluation, error detection, and accountability.

Key finding: The adaptive partnership scored 21 points above the human-only arrangement and 16 points above the AI-only arrangement, suggesting that complementary task allocation may create more value than simply choosing between human and machine judgment.

Source: Author-developed simulation. Values are illustrative and are not empirical research findings.

American Journal of Advanced Multidisciplinary Research and Innovation (AJAMRI)

E-ISSN: XXXX-XXXX ● Website: www.ajamri.com ● Email: editor@ajamri.com

6. Challenges

6.1 Automation Bias and Miscalibrated Trust

Researchers may accept an artificial intelligence recommendation because it is presented with numerical precision, technical complexity, or apparent confidence. This automation bias can become particularly strong when users lack the time or expertise required to independently reproduce the analysis. Conversely, some experts may reject valid machine findings because they conflict with established expectations or threaten professional autonomy.

Partnership design must therefore support calibrated rather than maximal trust. Confidence indicators should be empirically validated, limitations should be visible, and researchers should be able to inspect alternative explanations. Training should include examples of both successful and failed machine recommendations so that users understand the system's actual boundaries.

6.2 Explainability and Epistemic Validity

An explanation does not automatically provide scientific understanding. Feature attribution may identify which variables influenced a prediction without establishing causation. A visualization may appear intuitive while hiding sensitivity to model parameters. Scientific users must distinguish predictive explanation, mechanistic explanation, and causal explanation.

Explanation tools should therefore be evaluated for stability, fidelity, relevance, and comprehensibility. Where a high-stakes scientific claim depends on a particular mechanism, researchers may need an interpretable model, experimental intervention, or causal design rather than a post hoc explanation of an opaque predictor.

6.3 Data Quality, Provenance, and Reproducibility

Artificial intelligence can process poor-quality data efficiently and thereby create misleading confidence. Biased sampling, inconsistent measurement, missing metadata, and duplicated observations can undermine the resulting conclusion. Models trained on published literature may also inherit publication bias and disciplinary imbalance.

Every machine-assisted research workflow should preserve data provenance, preprocessing decisions, software versions, model parameters, prompts where applicable, random seeds, and human modifications. Without these records, other researchers cannot reconstruct how a conclusion was produced. Reproducibility becomes more difficult when systems are proprietary, frequently updated, or nondeterministic.

6.4 Responsibility and Authorship

Artificial intelligence cannot assume moral or scholarly responsibility for a research claim. Human authors must verify cited sources, validate calculations, examine generated content, and accept accountability for the final manuscript. Describing artificial intelligence as an author can obscure responsibility because the system cannot consent to publication, address conflicts of interest, or respond to allegations of misconduct.

Research organizations should define who may authorize AI-assisted conclusions, how disagreements are recorded, and when independent review is mandatory. Responsibility should correspond to actual control over study design, data interpretation, and publication decisions.

7. Discussion

7.1 Designing Complementary Research Intelligence

The simulated results support a complementary-intelligence model. Artificial intelligence contributes breadth, speed, consistency, and sensitivity to complex patterns. Researchers contribute theoretical understanding, contextual interpretation, ethical judgment, and responsibility. These capabilities should be integrated through deliberately designed interaction rather than combined informally.

Effective partnership begins with task decomposition. Literature screening, candidate ranking, image segmentation, anomaly detection, and repeated model comparison may be machine-led. Theory construction, construct definition, causal interpretation, ethical assessment, and final claim formulation should remain human-led. Some activities, including hypothesis refinement and model diagnosis, require repeated interaction.

The design should also preserve productive friction. A system that always confirms the researcher provides little protection against human bias, while a researcher who always approves the model provides little protection against machine error. Disagreement should trigger a documented process involving alternative models, additional evidence, sensitivity analysis, or independent review.

7.2 Explanation, Uncertainty, and Decision Authority

Explanations should be tailored to the scientific question. A chemist may require molecular substructures, a clinician may require comparable cases and uncertainty, and an environmental scientist may require spatial or temporal drivers. Generic confidence scores are insufficient where users need to evaluate whether a model has learned a scientifically plausible relationship.

Uncertainty should directly influence decision authority. A system operating within a familiar data region may be permitted to complete low-consequence analytical tasks automatically. When inputs are novel, uncertainty is high, or outcomes are consequential, the workflow should require human assessment. Adaptive authority is therefore more appropriate than a fixed rule assigning all final decisions to either humans or machines.

Human authority must nevertheless be substantive. A researcher who lacks time, information, or institutional permission to challenge the system does not provide meaningful oversight. Interface design, workload, training, organizational culture, and escalation mechanisms all determine whether human control functions in practice.

7.3 Implications for Scientific Institutions

Research institutions should establish policies for disclosure, validation, and documentation of artificial intelligence use. These policies should distinguish routine assistance from substantive analytical involvement. Automated grammar correction requires a different level of scrutiny from machine-generated hypotheses, statistical analysis, or interpretation.

Peer review may also need to examine the role of artificial intelligence in producing results. Authors should disclose material computational assistance, provide sufficient methodological detail, and verify references. Editors and reviewers should focus on evidential quality rather than assuming that the presence or absence of AI determines scientific merit.

Training programs should prepare scientists to work critically with intelligent tools. Relevant competencies include data literacy, model evaluation, uncertainty interpretation, bias detection, reproducible computation, research ethics, and the ability to recognize when a question requires causal or experimental evidence rather than predictive modeling.

8. Conclusion

Artificial intelligence can significantly extend scientific analytical capacity, but it does not eliminate the need for human expertise. Scientific decisions depend on theoretical meaning, contextual validity, causal interpretation, methodological appropriateness, ethical judgment, and responsibility. These functions cannot be reduced to predictive performance alone.

The simulation-based comparison found that combined decision arrangements performed more favorably than isolated human or machine processes. Human review improved the value of machine recommendations, while explanations supported more informed evaluation. The adaptive partnership achieved the strongest simulated result because it allocated tasks according to capability, uncertainty, novelty, and consequence.

The analysis also shows that collaboration can fail when trust is poorly calibrated, explanations are misleading, provenance is lost, or human oversight becomes ceremonial. Effective partnerships require transparent evidence, independent reasoning, structured disagreement, uncertainty-sensitive escalation, and complete documentation. Human researchers must retain accountability for scientific claims even when artificial intelligence contributes substantially to their development.

The principal limitation is that the numerical results are illustrative. Empirical validation should compare partnership models across scientific disciplines, levels of expertise, decision complexity, and consequences of error. Future studies should measure accuracy, time, reproducibility, calibration, disagreement resolution, cognitive load, and the distribution of benefits across researchers and institutions. The broader conclusion remains that artificial intelligence is most valuable not as a substitute for scientific judgment, but as a carefully governed partner that expands what researchers can examine while preserving the human responsibility to determine what the evidence means.

References

1. Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine, 35*(4), 105–120. https://doi.org/10.1609/aimag.v35i4.2513
2. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). https://doi.org/10.1145/3290605.3300233

3. Bansal, G., Nushi, B., Kamar, E., Weld, D. S., Lasecki, W. S., & Horvitz, E. (2019). Updates in human-AI teams: Understanding and addressing the performance/compatibility tradeoff. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(1), 2429–2437. https://doi.org/10.1609/aaai.v33i01.33012429
4. Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018). "It's reducing a human being to a percentage": Perceptions of justice in algorithmic decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). https://doi.org/10.1145/3173574.3173951
5. Boobier, S., Osbourn, A., & Mitchell, J. B. O. (2017). Can human experts predict solubility better than computers? *Journal of Cheminformatics, 9*, 63. https://doi.org/10.1186/s13321-017-0250-y
6. Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), 1–21. https://doi.org/10.1145/3449287
7. Cai, C. J., Winter, S., Steiner, D., Wilcox, L., & Terry, M. (2019). "Hello AI": Uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. *Proceedings of the ACM on Human-Computer Interaction, 3*(CSCW), 1–24. https://doi.org/10.1145/3359206
8. Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126. https://doi.org/10.1037/xge0000033
9. Gil, Y., Greaves, M., Hendler, J., & Hirsh, H. (2014). Amplify scientific discovery with artificial intelligence. *Science, 346*(6206), 171–172. https://doi.org/10.1126/science.1259439
10. Holzinger, A. (2016). Interactive machine learning for health informatics: When do we need the human-in-the-loop? *Brain Informatics, 3*, 119–131. https://doi.org/10.1007/s40708-016-0042-6
11. Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons, 61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007
12. Jiménez-Luna, J., Grisoni, F., & Schneider, G. (2020). Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence, 2*, 573–584. https://doi.org/10.1038/s42256-020-00236-4
13. Kamar, E. (2016). Directions in hybrid intelligence: Complementing AI systems with human intelligence. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (pp. 4070–4073).
14. Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics, 133*(1), 237–293. https://doi.org/10.1093/qje/qjx032
15. Kutchukian, P. S., Vasilyeva, N. Y., Xu, J., Lindvall, M. K., Dillon, M. P., Glick, M., Coley, J. D., & Brooijmans, N. (2012). Inside the mind of a medicinal chemist: The role of human bias in compound prioritization during drug discovery. *PLOS ONE, 7*(11), e48476. https://doi.org/10.1371/journal.pone.0048476
16. Lai, V., & Tan, C. (2019). On human predictions with explanations and predictions of machine learning models: A case study on deception detection. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 29–38). https://doi.org/10.1145/3287560.3287590

17. Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics, HFE-1*(1), 4–11. https://doi.org/10.1109/THFE2.1960.4503259
18. Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005
19. Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence, 267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007
20. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*, 206–215. https://doi.org/10.1038/s42256-019-0048-x
21. Schramowski, P., Stammer, W., Teso, S., Brugger, A., Herbert, F., Shao, X., Luigs, H.-G., Mahlein, A.-K., & Kersting, K. (2020). Making deep neural networks right for the right scientific reasons by interacting with their explanations. *Nature Machine Intelligence, 2*, 476–486. https://doi.org/10.1038/s42256-020-0212-3
22. Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe and trustworthy. *International Journal of Human–Computer Interaction, 36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118
23. Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019). Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–15). https://doi.org/10.1145/3290605.3300831