Journal of Research and Innovation in Technology, Commerce and Management
Vol. 3 Issue 9, September 2026, pp. 39099-39107
ISSN: 3049-3129 (Online)

GENERATIVE AI-DRIVEN SCIENTIFIC WORKFLOWS: A PROVENANCE-CENTERED FRAMEWORK FOR REPRODUCIBLE AND SCALABLE DISCOVERY

ISHRAT TAMREEN
Dept. of Computer Science & Engineering
Lords Institute of Engineering and Technology, Hyderabad
Email: ishrattamreen@gmail.com

AMRAH MOHAMMADI
Dept. of Computer Science & Engineering
Lords Institute of Engineering and Technology, Hyderabad
Email: mohammadiamrah26@gmail.com

SOBIYA KHAN
Dept. of Computer Science & Engineering
Lords Institute of Engineering and Technology, Hyderabad
Email: sobiyasultana18@gmail.com

Cite as: ISHRAT TAMREEN, AMRAH MOHAMMADI, & SOBIYA KHAN. (2026). GENERATIVE AI-DRIVEN SCIENTIFIC WORKFLOWS: A PROVENANCE-CENTERED FRAMEWORK FOR REPRODUCIBLE AND SCALABLE DISCOVERY. Journal of Research and Innovation in Technology, Commerce and Management, Vol. 3(Issue 9), 39099–39107. https://doi.org/10.5281/zenodo.22806586

DOI: https://doi.org/10.5281/zenodo.22806586

ABSTRACT—Generative AI is widely used in scientific research to assist with literature reviews, hypothesis formation, data analysis, and experimentation. While these capabilities revolutionize discovery, they present reproducibility challenges due to undocumented workflows and human decisions. To address this, we introduce TRACE-Sci, a provenance-based workflow for AI-assisted research. TRACE-Sci treats computational artifacts as auditable objects, linking them to evidence, execution context, and validation results through five components: task identification, evidence anchoring, tool execution, context locking, and risk-based checkpoints. We assessed the architecture using a Monte Carlo simulation across 120 synthetic workflow instances. Compared to direct and tool-augmented LLM workflows, TRACE-Sci achieved a mean provenance completeness of 91 percent, a rerun success rate of 88 percent, and a citation verification rate of 93 percent. The breakdown demonstrates that provenance completeness increases non-linearly with human review coverage. These results suggest that scientific AI systems should be evaluated based on their auditability and reproducibility rather than just

task completion.

Keywords—Generative Artificial Intelligence; Scientific Workflows; Reproducibility; Provenance; Large Language Models; Research Agents; Human-in-the-loop; Computational ScienceI.

INTRODUCTION

Scientific research nowadays relies more and more on computational workflows which involve literature searches, structured data, simulation, statistical analysis, software execution, visualization, and scientific reporting. Generative AI has now introduced an additional stage to these workflows. Large language models are capable of turning research objectives into plans, producing code and queries, summarizing evidence, interpreting intermediate results, and suggesting possible hypotheses. There have been recent studies showing the development of systems that carry out various stages of the research process, such as literature review, experimentation, coding, and report generation, and other research has looked at the use of language agents in scientific and laboratory environments [1]–[4]. The key issue, then, is not merely whether these systems are able to carry out individual research tasks. The more complicated question is whether their outputs can be integrated into scientific workflows without damaging traceability and reproducibility. A scientific result has no value unless researchers can find out where it came from, which evidence backed it, what transformations were carried out, and under what computational conditions the result was obtained. Traditional computational workflow systems have already dealt with a number of the issues involved. By means of explicitly defined dependencies, environments that are versioned, burnable pipelines, and records of provenance it is possible to reconstruct many computational processes. For instance, Nextflow was developed in order to enable reproducible data-intensive workflows by explicitly managing the computational processes and their environments [5]. The W3C PROV model offers a formal vocabulary for describing entities, activities, agents, and the relationships between them [6]. AI-assisted scientific workflows involve new kinds of state which are less frequently found in conventional workflow systems, such as different versions of models, the stochastic behavior of models, various prompt configurations, different retrieval results, varying tool-selection decisions, and changes in external services. If these factors are not recorded, then two runs of what seems to be the same workflow may yield different results without offering a clear explanation for the difference.

The paper mainly treats the problem as one of workflow engineering and provenance rather than merely asking about the capabilities of the model. We have developed TRACE-Sci, an architecture focused on provenance which is intended to work in conjunction with existing language models and scientific tools. The architecture is based on five principles: 1. stable task identifiers; 2. evidence-linked generation; 3. execution supported by tools and capable of being audited; 4. and the explicit recording of context; 5. validation checkpoints based on the risk associated with the workflow. The main point is that a scientific AI workflow should preserve not only its final result but also the information needed to understand it and, if possible, to reproduce the process by which that result was obtained. We use a controlled Monte Carlo simulation to examine this idea. It should be understood that the simulation is not an empirical assessment of a scientific system that is already in operation; rather, it offers a controlled setting in which various workflow architectures can be compared, all based on the same assumptions. The study is concerned with whether explicit provenance and verification mechanisms enhance traceability and the behavior of the model when it is rerun, and also with how these properties change as the coverage of human checkpoints varies.

II. RELATED WORK

The study of AI in scientific contexts includes a wide variety of tasks, such as literature reviews, the generation of hypotheses, experimental design, data interpretation, and automated discovery. Recent reviews have highlighted both

the potential of large language models throughout the scientific method and the necessity of evaluation criteria that are based on scientific goals rather than just the quality of language generation [1]. Other surveys have also pointed out long-horizon scientific agents, domain-specific representations, and integrated reasoning and tool use as key areas for future research [2]. Recent systems have shown how research is moving from providing conversational assistance to involving multi-stage scientific workflows. For instance, Agent Laboratory combines literature reviews, experimentation, and the preparation of reports into a research-assistant framework [3]. In another case, a system based on a large language model presented a workflow that progressed from annotated data through hypothesis generation and the production of analysis code to the final results and to manuscripts that can be verified by humans [4]. While these systems prove that it is possible to use language models at various stages of research, they also point out a major concern regarding reproducibility: the final coherent document does not always reveal all the intermediate decisions or dependencies that had an influence on the result. Recent benchmarks have provided additional evidence showing that scientific replication is much more difficult than generating outputs that seem plausible in a local context. PaperBench assesses whether AI agents are able to reproduce published machine-learning research from scratch and achieved a best-tested average replication score of 21.0% [7]. ScienceAgentBench tests language agents on validated tasks taken from scientific publications, including the generated programs, the execution results, and the associated costs [8]. Evaluations focused on laboratory settings have also started to look at AI agents in situations that involve experimental design, interaction with instruments, and the analysis of results [9]. The fact that these developments have taken place shows that when assessing scientific agents the evaluation should go beyond simply looking at the quality of the text produced. While the ability to carry out a task is important, it is just as important to be able to identify the evidence used, to preserve the relevant computational state, to recover from a failure, and to reproduce the important artifacts. The necessary foundations for such an approach are already available in the field of reproducible computing and in research data infrastructure. According to the FAIR principles, research objects should be findable, accessible, interoperable, and reusable [10]. W3C PROV offers a formal method for describing how entities and activities are related and how outputs are derived [6]. Workflow systems like Nextflow show the advantages of using explicit computational environments and dependencies, while continuous-analysis methods have investigated mechanisms for automating reproducibility checks [5], [11]. While TRACE-Sci is based on these existing ideas it extends the provenance boundary to include AI-mediated operations, considering model configurations, retrieval contexts, prompts, tool calls, validation outcomes, and human decisions as workflow objects rather than as merely incidental execution details.

III. THE TRACE-SCI WORKFLOW ARCHITECTURE

A. Evidence Anchoring

The initial step is to create evidence anchors. Instead of generating the prose without support and then trying to add the citations later on, TRACE-Sci creates the evidence objects before reaching the relevant generation stage An evidence object includes a source identifier, a retrieval timestamp, the source location, and normalized textual or structured content. Claims that are produced as part of the workflow are given individual identifiers and can be categorized based on how they relate to the stored evidence, for example by being stated as supported, contradicted, uncertain, or derived. This representation is in keeping with the general aim of FAIR research objects and with provenance models which explicitly represent derivation relationships [6], [10].

TRACE-Sci is not a new foundation model but an orchestration architecture; its name comes from five elements: Task Registry, Retrieval and evidence anchoring, Auditable execution, Context locking, and Evaluation checkpoints. The process starts when a research objective is given a stable

task identifier. Once this has been done, evidence retrieval creates a source set and records the identifiers and the times taken to retrieve them. A planning module then turns the research objective into a structured workflow plan. However, the plan is not regarded as scientific evidence. The process then moves on to execution supported by tools. The explicit inputs are converted into code, data products, simulations, analyses, or other computational artifacts. Validation gates are used at the relevant stages. Once an artifact has been accepted it is linked both to the provenance ledger and to an immutable run bundle which includes the information needed to reconstruct the execution context.

The aim is to ensure that the workflow can be readily transferred and verified in different computational environments. In this approach, the unit of reproducibility is not just the final manuscript or script; rather, it is the entire recorded run together with its related artifacts.

B. Context Locking

The outcomes of scientific workflows can vary even if the official research task does not change; for example, a different model version, retrieval corpus, dependency version, or external service can lead to a different result. TRACE-Sci thus records the relevant execution context, comprising the name and version of the model, the system configuration, the hash of the prompt template, the retrieval configuration, the versions of the tools, the package-lock information, the container identifiers where applicable, the random seeds, and the execution timestamps. Context locking is not meant to ensure complete determinism since some scientific calculations and AI systems will still be stochastic or rely on external resources; rather, the aim is to make the conditions under which a result was obtained explicit so that later executions can be compared with the original run

C. Validation Gates

The conditions which are defined at each stage must be met before the execution can go ahead, and the right kind of checks will depend on the type of activity. For instance, the literature-retrieval stage might involve citation-resolution checks. The software-generation stage might include syntax checks and unit tests. The data-processing stage might require schema and range validation. In the results stage, it is necessary to verify that the reported metrics are consistent with the underlying computational outputs. Human inspection is set off by high-risk events not because it is required for each individual operation. Examples are unsupported claims, failed statistical assumptions, conflicting evidence, or the execution of a tool with low confidence. The design shifts the role of human oversight from one of continuous supervision to that of risk-based intervention.

D. Persistent Run Bundles

The complete workflow state is stored in a run bundle by TRACE-Sci. This bundle includes the workflow specification, the input manifest, the evidence manifest, the model and tool metadata, the execution logs, the intermediate artifacts, the validation results, and the final outputs.

IV. EVALUATION METHODOLOGY

A. Synthetic Benchmark

The benchmark that is simulated includes 120 workflow instances for each configuration, and each instance is shown as an eight-stage dependency graph comprising: 1. evidence retrieval; 2. planning; 3. the generation of code or tools; 4. execution; 5. data transformation; 6. analysis; 7. result checking; and8. reporting. The levels of task difficulty are three: low, medium, and high. The probabilities of failing are separately varied for retrieval, coding, execution, and validation. The following three workflow configurations are compared. Direct LLM workflow: The model generates and interprets the outputs while having only a limited amount of structured state and minimal persistent provenance. In a workflow that incorporates tools: the model is linked to retrieval and execution tools but does not need a persistent provenance graph. The workflow includes evidence anchoring, context locking, persistent

provenance, and risk-based human checkpoints. The three different configurations are based on the same synthetic task population.

This paper presents a framework and does not contain results from a finished laboratory experiment or a production-scale scientific deployment; the quantitative evaluation has therefore been carried out using a controlled Monte Carlo simulation. This distinction is important. The numerical results given in this section are synthetic outputs produced based on predefined assumptions They have not been obtained from measurements made on human participants, from physical experiments, or from actual scientific workloads. The aim of the simulation is to compare different workflow designs under controlled conditions and to look at how changes in verification coverage influence the proposed architecture.

TABLE I. SIMULATED WORKFLOW PERFORMANCE

Metric	Direct LLM	Tool-augmented	TRACE-Sci
Provenance completeness	52	73	91
Rerun success	58	76	88
Citation verification	61	79	93
Execution failure	19	11	6

Table I indicates the expected direction of improvement rather than a claim about an observed system. TRACE-Sci produces the highest simulated provenance and citation integrity while reducing modeled execution failures. The time index suggests an approximately 20 percent overhead reduction relative to the normalized direct workflow after accounting for fewer failed reruns. The interpretation is deliberately narrower than a claim of faster real-world research: in practical deployments, logging and verification can increase wall-clock time even when they reduce rework.

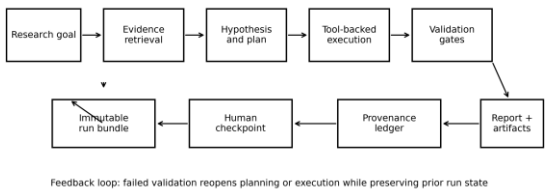


Fig. 1. TRACE-Sci architecture.

The persistent provenance ledger and immutable run bundle surround the workflow so that validation failures can be corrected without discarding prior state. C. The interpretation of the simulation The simulation must be regarded as a scenario-based assessment of the architecture, not as proof of its performance in an actual scientific environment. Specifically, it is the numerical assumptions that give rise to the absolute figures of the reported metrics. Different choices of parameters could lead to different results. Accordingly, the reason for carrying out the simulation is to make the assumptions and the comparative behavior clear, not to suggest that TRACE-Sci has in fact reached the reported percentages.

B. Evaluation Metrics

Five metrics are modeled. The degree to which the provenance is complete refers to the percentage of workflow objects that are recorded and linked to the final result. The success of a rerun refers to the ability of a second execution, using the stored configuration, to reproduce the expected class of artifact without the need to manually reconstruct the hidden state. The process of verifying citations calculates the percentage of claimed references that can be linked to stored evidence objects. The failure rate of executions refers to the proportion of workflow instances that end as a result of an uncorrected computational or orchestration error. For the direct LLM workflow, the time-to-result is set at 100. This figure is meant to reflect the balance that has been modeled between the extra logging and validation and the decrease in rework resulting from failed executions. The values given are the averages obtained from the 120 simulated instances of the workflow and the

uncertainty intervals are calculated from the Monte Carlo samples.

V. SIMULATED RESULTS AND ANALYSIS

The simulation shows that the main benefit of TRACE-Sci is better traceability not an increase in the amount of tasks completed. With the assumptions used in the model, provenance completeness rises from 52% in the case of the direct workflow to 73% in the case of the tool-augmented workflow and reaches 91% with TRACE-Sci. This difference is due to the greater amount of workflow state made explicit by each architecture, since tool integration causes further execution events to be recorded and TRACE-Sci additionally records the relationships between evidence, model context, intermediate artifacts, claims, and validation outcomes. The effect is especially important in the case of multi-stage workflows; a final result can be the outcome of a number of transformations, and it may be difficult to reconstruct those transformations solely from the final report. A persistent provenance graph gives an explicit representation of those dependencies. In the TRACE-Sci simulation citation verification reaches 93 percent. This figure must be understood in a narrow sense since it shows that, on the basis of the assumptions of the simulation, the procedure is better suited to determining whether a cited claim matches a stored evidence object and whether that evidence was available to the generation process. It does not prove that the original scientific claim is true. It is important to make this distinction since citation verification and scientific validity are two separate features; a claim may refer to a true and relevant source and yet still arrive at an incorrect conclusion based on that source. In the TRACE-Sci configuration the rate of successful reruns is 88%, which is less than the level of provenance completeness that had been modeled. This discrepancy shows an important limitation of provenance systems: just because a state is recorded does not mean that the recorded state is enough to reproduce every result. External dependencies could vanish or alter. Remote services might return different results. Numerical libraries might show different behavior in different hardware or software environments. Scientific datasets too may be updated after the original run. That is why reproducibility is more appropriately seen as a gradable characteristic than as a simple yes-or-no category. A workflow should make a distinction between artifacts that have been reproduced exactly, those that have been reconstructed under equivalent conditions, and outputs that still depend on external resources. The simulation thus leads to a more conservative interpretation of reproducibility: provenance improves the conditions under which reproduction can take place, but provenance by itself is not sufficient to ensure reproduction.

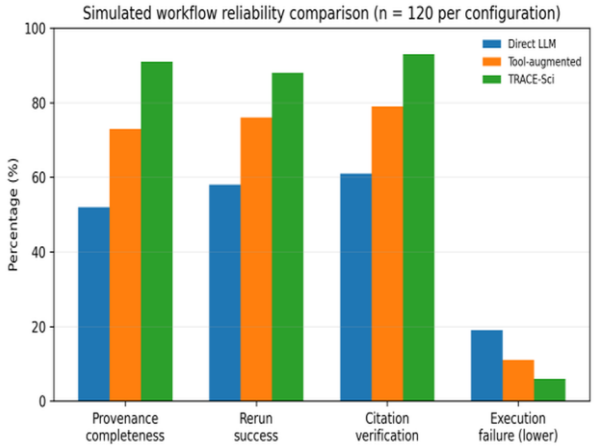


Fig. 2. Simulated comparison of three workflow designs across traceability, rerun behavior, citation verification, and execution failure.

Results are synthetic model outputs, not empirical observations.

VI. SENSITIVITY AND SCALABILITY

A key question is the extent to which human supervision is necessary in the case of a provenance-aware scientific workflow. A full manual review would negate the advantages of automation and might become impractical when applied to long-running workflows; TRACE-Sci therefore calculates human checkpoint coverage as the percentage of workflow transitions which are subject to human verification. There is a nonlinear relationship between the level of checkpoint coverage and provenance

completeness produced by the simulation, with the greatest improvement in the model occurring in the range from zero to about 60% checkpoint coverage and only smaller gains being obtained beyond that range. The result indicates that a risk-based approach to review is advisable, since high-impact or uncertain transitions should be subject to more extensive human oversight whereas routine and low-risk transformations can continue to be automated.

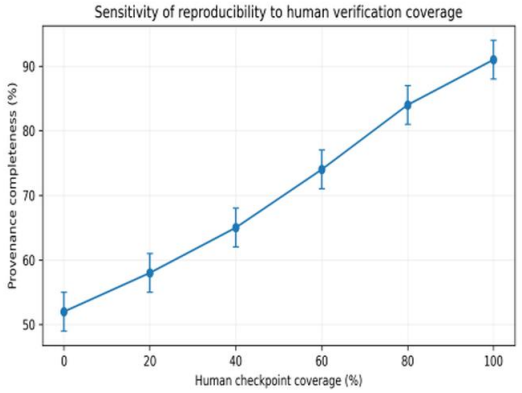


Fig. 3. Simulated sensitivity of provenance completeness to human checkpoint coverage.

A. Provenance Storage

Scalability also presents a problem when it comes to storage; if every model response is recorded in full, this could greatly increase the size of the provenance record, especially in the case of workflows that involve long contexts or large computational artifacts. To deal with this, TRACE-Sci separates the immutable artifacts from the compact provenance metadata so that the large artifacts can stay in a content-addressed storage system while the provenance graph keeps track of the identifiers, hashes, and relationships. This idea is in keeping with reproducibility infrastructures in which the execution of workflows and the storage of artifacts are regarded as related but separate issues [5], [6].

Error bars represent the modeled 95 percent simulation interval. It should still be regarded as a feature of the simulation and not as a general rule. The degree to which human inspection is appropriate will vary according to the scientific field in question, the consequences of a failure, and the reliability of the tools used.

VII. DISCUSSION

The idea is that generative AI should be included in scientific procedures through clearly defined limits, rather than being regarded as a black-box, general-purpose assistant. A language model is just one part of the research process; a scientific agent may also make use of evidence repositories, datasets, software environments, statistical tools, laboratory interfaces, and human reviewers. Hence, assessing only the generated text gives an incomplete view of the system's reliability. Another point is what is meant by autonomy. TRACE-Sci doesn't aim at getting rid of human involvement; rather, it distinguishes operational autonomy from scientific accountability. It is possible for a workflow to include several stages without requiring direct human involvement and yet still ensure that the information necessary for a researcher to examine key decisions is retained. In this approach, human oversight is of a supervisory nature rather than being continuous. This method is in agreement with research which indicates that human participation can enhance the quality of AI-assisted research processes [3]. Another point to consider is how the system handles failure. In ordinary software systems, logs are usually regarded mainly as tools for debugging. Whereas in scientific workflows the execution records can also count as part of the scientific evidence. For instance, altering the datasets chosen, the retrieval results, the version of the dependency, or the numerical output might influence the final scientific conclusion; it is for this reason that such events should be stored in a format that allows for retrospective analysis. Therefore, TRACE-Sci regards provenance as a subject of research and not merely as an operational log. The framework also has obvious limitations. First of all, even if the provenance is available the scientific reasoning may still be flawed. A workflow that has been accurately recorded can nevertheless lead to an invalid scientific conclusion. Second, it is not possible to ensure that a complete run bundle will be able to reproduce the results when the workflow relies

on external services, physical instruments, unavailable datasets, or on changing computational infrastructure. Thirdly, validation rules are inherently dependent on the particular domain; general checks are capable of detecting syntax, schema, or consistency errors but cannot take the place of expert judgment on matters of causal inference, clinical significance, experimental feasibility, or theoretical validity. Fourth, the quantitative assessment is synthetic. Although the simulation offers a controlled method of examining the proposed architecture, it does not prove that TRACE-Sci achieves the same level of performance when used in real scientific workloads. They ought to be seen as requirements for future empirical evaluation rather than as characteristics which can be settled merely by the provenance layer.

VIII. THREATS TO VALIDITY

The main issue relating to internal validity is the selection of the simulation parameters. The values assigned to the three workflow configurations are based on plausible scenarios not on empirical measurements. Different assumptions could lead to changes in both the absolute results and in the differences between the configurations. In order to address this problem, the simulation applies the same synthetic workflow population to all three configurations and clearly states that the results are modeller. Another issue is construct validity. Although provenance completeness, rerun success, and citation verification are useful in indicating that a workflow is reproducible, none of them is a full measure of scientific validity. Another issue is external validity. Scientific workflows vary greatly depending on the domain, the type of data, the equipment used, the computational demands, and the regulatory environment. A result derived from a synthetic workflow graph cannot be directly applied to laboratory science or to high-stakes research situations. In the end, the simulation assesses the process properties rather than the scientific value of the discoveries made by the workflows. It does not judge the novelty, correctness, or experimental significance of the scientific hypotheses generated by the workflows.

IX. CONCLUSION AND FUTURE RESEARCH

Generative AI is currently becoming capable of participating in scientific workflows that are becoming increasingly intricate. The challenge involved in attaining calculable use in this field is thus not merely a question of enhancing language generation; it also necessitates mechanisms which maintain evidence, computational context, intermediate artifacts, the outcomes of validation, and human decisions. The article introduces TRACE-Sci, a framework based on provenance which integrates evidence anchoring, context locking, verifiable tool execution, validation gates, and persistent run bundles. A Monte Carlo simulation was performed in order to compare TRACE-Sci with both a direct LLM strategy and a tool-augmented strategy. Under the circumstances assumed in the simulation, TRACE-Sci attained a higher level of modeled provenance completeness, a superior rate of rerun success, and better citation verification, while simultaneously having a lower modeled execution-failure rate. However, these findings are *artificial and methodological* and must not be regarded as empirical measurements or as evidence of the performance of the strategy in a real scientific system. The subsequent phase involves conducting an empirical assessment on real scientific workloads. Future research should analyze the architecture on tasks such as scientific code generation, literature-based analysis, and the replication of published computational studies. These evaluations should compare several current models under controlled conditions regarding both the tools and the environment utilized, and should include reports on computational cost, latency, failure recovery, and the effort needed by reviewers. It is especially vital to assess whether workflows that are aware of provenance decrease substantive scientific errors, rather than just improving the recording of workflow events. Future evaluations should therefore involve domain experts, with the goal of testing whether improved provenance actually leads to a better ability to detect incorrect assumptions, unsupported claims, and irreducible results. The objective is not to attain full autonomy for scientific research. Instead, it is

to develop scientific agents which can perform routine and computationally demanding tasks on a massive scale, while leaving a record that is scrutinized and reusable which links the research query to the evidence, the method of execution and the conclusions reached.

ACKNOWLEDGMENT

The authors acknowledge the use of publicly available scientific literature and established provenance and workflow concepts as the basis for this framework. No participant data, patient data, or private institutional dataset was used in the simulation reported in this manuscript.

REFERENCES

[1] Y. Zhang, S. A. Khan, A. Mahmud, H. Yang, A. Lavin, M. Levin, et al., "Exploring the role of large language models in the scientific method: from hypothesis to discovery," npj Artificial Intelligence, vol. 1, Art. no. 14, 2025, doi: 10.1038/s44387-025-00019-5.

[2] C. K. Reddy and P. Shojaee, "Towards scientific discovery with generative AI: progress, opportunities, and challenges," Proc. AAAI Conf. Artificial Intelligence, vol. 39, no. 27, pp. 28601–28609, 2025, doi: 10.1609/aaai.v39i27.35084.

[3] S. Schmidgall et al., "Agent Laboratory: Using LLM agents as research assistants," in Findings of the Association for Computational Linguistics: EMNLP 2025, pp. 5977–6043, 2025, doi: 10.18653/v1/2025.findings-emnlp.320.

[4] T. Ifargan, L. Hafner, M. Kern, O. Alcalay, and R. Kishony, "Autonomous LLM-driven research—from data to human-verifiable research papers," NEJM AI, vol. 2, no. 1, 2025, doi: 10.1056/AIoa2400555.

[5] P. Di Tommaso, M. Chatzou, E. W. Floden, P. Prieto Barja, E. Palumbo, and C. Notredame, "Nextflow enables reproducible computational workflows," Nature Biotechnology, vol. 35, pp. 316–319, 2017, doi: 10.1038/nbt.3820.

[6] W3C, "PROV Model Primer," W3C Working Group Note, Apr. 30, 2013.

[7] G. Starace et al., "PaperBench: Evaluating AI's Ability to Replicate AI Research," in Proc. 42nd Int. Conf. Machine Learning, vol. 267, pp. 56843–56873, 2025.

[8] Z. Chen et al., "ScienceAgentBench: Toward Rigorous Assessment of Language Agents for Data-Driven Scientific Discovery," in Proc. Int. Conf. Learning Representations, 2025.

[9] I. Mandal et al., "Evaluating large language model agents for automation of atomic force microscopy," Nature Communications, vol. 16, Art. no. 9104, 2025, doi: 10.1038/s41467-025-64105-7.

[10] M. D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship," Scientific Data, vol. 3, Art. no. 160018, 2016, doi: 10.1038/sdata.2016.18.

[11] B. K. Beaulieu-Jones and C. S. Greene, "Reproducibility of computational workflows is automated using continuous analysis," Nature Biotechnology, vol. 35, no. 4, pp. 342–346, 2017, doi: 10.1038/nbt.3780.