

## Argonne National Laboratory

### Argo Research Agent: Generative AI-assisted Scientific Discovery at Argonne

Argonne National Laboratory relies on secure AI systems to support operations and scientific research. To enhance these capabilities, the team extended "Argo," a generative AI tool, to support complex, evidence-based reporting. The team implemented and evaluated two distinct agentic architectures to optimize performance. The first approach utilized a Parallel Agent System, where specialized agents independently generated questions (question agent), executed searches (answer agent), and synthesized reports (reports agent). The second approach employed a ReAct agent that accessed tools directly through cyclic reasoning steps.

Argo produces clean, structured research reports with automatic in-text citations and a clearly organized "Sources" section. As shown in the preview, the system generates a topic header, a concise evidence-based summary, and embedded references that point to the source material used in the analysis. The interface highlights exactly where information comes from and automatically stores the generated report, giving users a transparent record of the model's reasoning process.

### Additional Features

Additional capabilities—such as short-term memory to stabilize multi-step reasoning and an observability layer that exposes tool use and intermediate steps—enhanced reliability across both architectures. These features increased consistency, traceability, and confidence in the resulting reports.

**Topic:** latest research about evaluating AI models

The latest research on evaluating AI models emphasizes comprehensive, multi-faceted methodologies that go beyond traditional metrics. Notably, the paper "EAIRA" (arXiv:2502.20309v1) introduces a holistic framework that includes factual recall tests, reasoning assessments, and real-world, iterative experiments such as lab-style and field-style evaluations. These approaches aim to accurately capture AI capabilities in supporting scientific research, addressing the complexity and domain-specific nature of scientific tasks. Recent trends highlight the importance of real-world interactions and domain-specific benchmarks to better understand AI strengths and weaknesses in scientific applications.

## Sources

[1] EAIRA: Establishing a Methodology for Evaluating AI Models as Scientific Research Assistants

<https://arxiv.org/abs/2502.20309v1>

❖ Report saved to:  
`data/reports/20251209_174251_latest_research_about_evaluating_AI_models.md`