Research Highlight | Open Access

AI Agent

Xing *et al. AI Agent* **2026**, *2*, 33 **DOI:**10.20517/aiagent.2026.27

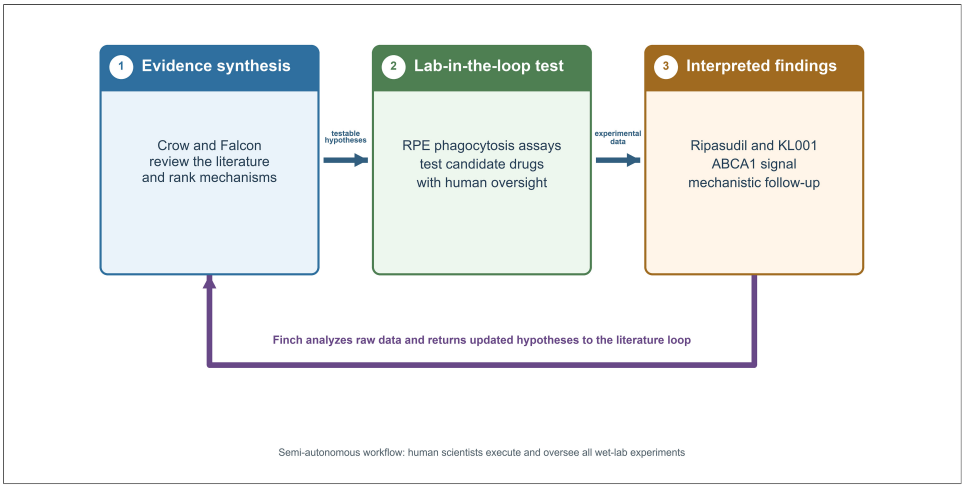
Robin as a lab-in-the-loop multi-agent system for scientific discovery

Xiujing Xing[1,#], Julie Ying Zhang[2,#], Jiaao Wang[3], Jeff Celaje[4]

Citation: Xing, X.; Zhang, J. Y.; Wang, J.; Celaje, J. Robin as a lab-in-the-loop multi-agent system for scientific discovery. *AI Agent* **2026**, *2*, 33. https://dx.doi.org/10.20517/aiagent.2026.27

Received: 17 Jun 2026 **First Decision:** 24 Aug 2026 **Revised:** 1 Sep 2026 **Accepted:** 17 Sep 2026 **Published:** 6 Nov 2026

Academic Editor: Aloysius Soon **Copy Editor:** Tong Wang **Production Editor:** Tong Wang



MAIN TEXT

Robin is best read as a careful lab-in-the-loop demonstration rather than as proof of fully autonomous biology. In the *Nature* study, Ghareeb *et al.* linked literature search, evidence synthesis, hypothesis ranking, experimental planning, computational analysis and hypothesis revision around one disease case: dry age-related macular degeneration (dAMD)[1]. Human scientists still chose the problem frame, performed the wet-lab work and retained responsibility for interpretation. The central message of the Highlight is therefore architectural: Robin makes the intellectual loop of discovery more explicit and auditable, but it does not remove the need for experimental validation.

This contribution should be placed in a broader landscape of agentic science. Earlier work had already explored retrieval-augmented ontological graphs and multi-agent materials-design workflows[2], autonomous chemical research agents that plan and execute laboratory tasks[3], end-to-end artificial intelligence (AI)-research systems

Check for updates

[1]Department of Chemistry, University of California, Davis, Davis, CA 95616, USA.
[2]College of Health and Human Sciences, San Jose State University, San Jose, CA 95192, USA.
[3]Department of Chemistry, The University of Texas, Austin, Austin, TX 78712, USA.
[4]Department of Chemistry, University of Southern California, Los Angeles, CA 90089, USA.
[#]These authors contributed equally to this work.

Correspondence to: Dr. Xiujing Xing, Department of Chemistry, University of California, Davis, Davis, CA 95616, USA. E-mail: xiuxing@ucdavis.edu

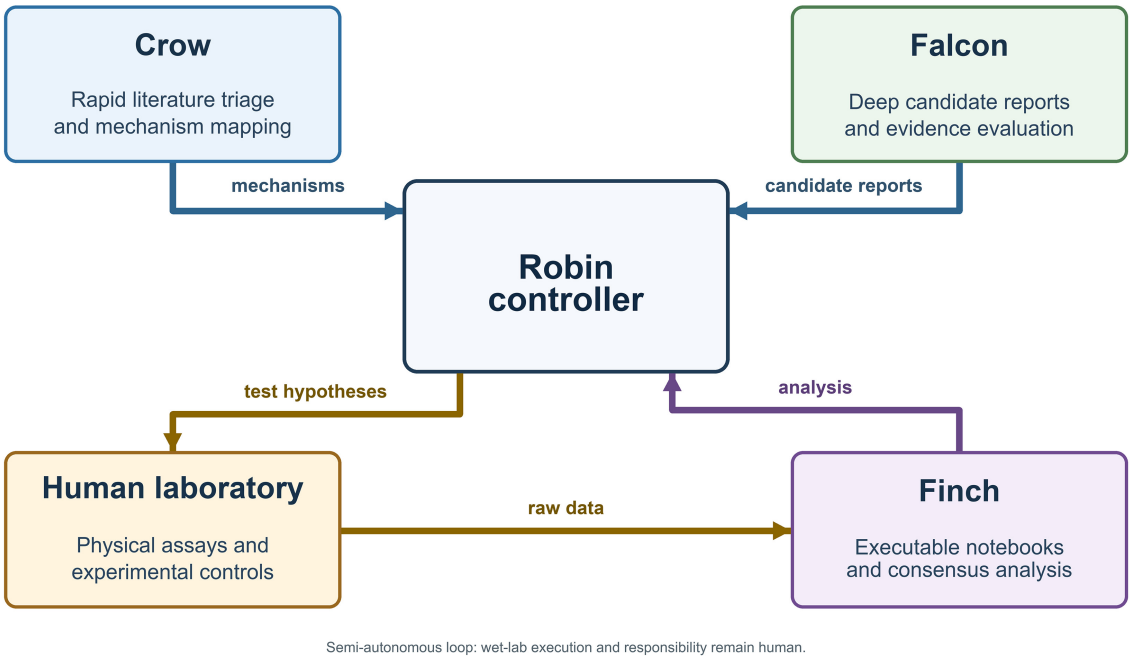


Figure 1. Robin's lab-in-the-loop architecture.

Table 1. Robin in relation to prior agentic-science systems

System or direction	What it demonstrates	How Robin is positioned
RAG and multi-agent materials design[2]	Graph-based retrieval, multi-agent reasoning and executable simulations	Shares the multi-agent/tool idea, but applies it to biomedical repurposing with wet-lab feedback
Coscientist/Autonomous chemistry[3]	LLM planning and orchestration for chemistry workflows	Robin emphasizes literature-grounded therapeutic hypotheses and post-experiment biological data analysis
AI Scientist and related AI-research agents[4]	End-to-end computational research loops, manuscript generation and automated review	Robin's experiments are biological and lab-in-the-loop, so physical validation remains human executed
ScienceClaw plus Infinite swarms[5]	Distributed agents coordinate through artifacts, provenance and shared scientific discourse	Robin is more centrally scaffolded around one disease workflow rather than open-ended swarm discovery

AI: Artificial intelligence; RAG: retrieval-augmented generation; LLM: large language model.

that generate ideas, run computational experiments and write papers[4], and decentralized scientific-agent swarms such as ScienceClaw plus Infinite[5]. Robin differs from these examples less by being the first multi-agent scientific system and more by joining source-grounded biomedical hypothesis generation with follow-up analysis of newly produced experimental data in a single therapeutics workflow. Table 1 summarizes these distinctions.

The system itself is deliberately modular [Figure 1]. Crow and Falcon are literature agents built on the PaperQA2 line of retrieval and evidence-synthesis work[6]: Crow provides fast orientation, whereas Falcon prepares deeper candidate reports. Finch writes and executes analysis code for data modalities such as flow cytometry and RNA sequencing. The division of labor is useful, but the ablation results should be interpreted with caution: better performance may arise from retrieval discipline, citation checking, tool access and workflow scaffolding, not from agent specialization alone.

Table 2. Evidence claims and interpretive cautions

Claim	Supported by	Important caution
RPE phagocytosis is a plausible dAMD intervention axis	Literature review, mechanism ranking and the selected phagocytosis assay[1]	One disease use case cannot prove general discovery performance
Ripasudil is a promising repurposing candidate	ARPE-19 and RPE-SC assays, dose-response trends and known ocular use[1]	*In vivo* models, dosing, durability and clinical endpoints remain untested
KL001 broadens the search space	Unexpected positive signal in RPE-SC screening[1]	Mechanism and reproducibility require follow-up work
Agent scaffolding improves reliability	Crow/Falcon ablations and Finch task evaluations[1]	The causal contribution of multi-agent decomposition is intertwined with retrieval, verification and tool use

RPE: Retinal pigment epithelium; SC: steam cell; dAMD: dry age-related macular degeneration; KL001: drug developed by Cayman Chemical Company; Cat No. 13878; ARPE-19 cells: ATCC; Cat. No. CRL-2302.

The dAMD case is compelling because Robin moved from a disease label to an experimentally testable phenotype. It reviewed 151 papers to nominate ten disease mechanisms, prioritized enhancement of retinal pigment epithelium (RPE) phagocytosis, and then reviewed about 400 additional papers to propose candidate compounds[1]. The initial positive result for Y-27632 supported the literature-derived reasoning but did not by itself constitute a therapeutic discovery; prior preclinical work had already shown that a clinically used rho kinase, Ripasudil (ROCK) inhibition can restore RPE phagocytic capacity in relevant contexts[7].

The stronger translational signal came in the second iteration. Robin nominated ten additional compounds, and Finch's analysis reported that ripasudil, a glaucoma-approved ROCK inhibitor, increased ARPE-19 phagocytosis 1.89-fold relative to dimethyl sulfoxide controls; human analysis showed a 1.75-fold increase, and dose-response work plus primary human RPE stem-cell-derived assays supported greater potency than Y-27632 (Drug developed by Cayman Chemical Company; Cat. No. 10005583)[1]. KL001, a circadian-clock modulator, also emerged as an unexpected RPE-steam cell (SC) hit[1]. These findings are intriguing but remain preclinical and *in vitro*, with false-positive risk, assay dependence and replication requirements still unresolved. Table 2 summarizes the evidence claims and interpretive cautions.

Robin's RNA-seq loop further shows how an agent can generate mechanistic follow-up rather than stop at a screening hit. After Y-27632 treatment, Finch identified a threefold critical lipid efflux pump (ABCA1) upregulation with a very small adjusted *P*-value[1]. ABCA1 is biologically plausible because lipid efflux in RPE is linked to retinal homeostasis and AMD-related degeneration[8]. Still, transcriptional association is not a mechanism: additional perturbation studies would be needed to prove that ABCA1 mediates the phagocytosis phenotype. Please note that this Research Highlight does not add new experiments; it clarifies what the reported study does and does not establish.

The comparison with OpenAI Deep Research is informative but not decisive. Deep Research generated 17 unique candidates that were screened in RPE-SCs and did not identify ROCK inhibition[1]. However, retrieval strategy, prompt framing, tool access, candidate-selection policy and experimental context may all have differed. The fairest conclusion is not that one product is categorically superior, but that domain-specific scaffolds can matter when the scientific objective is narrow, evidence-sensitive and experimentally constrained.

Robin therefore marks a meaningful step toward AI-organized discovery, while also illustrating the new evaluation bottleneck. Future systems will need transparent provenance, calibrated novelty checks, negative-result handling, reproducible notebooks, expert review and safety governance before they can be trusted

across diseases or disciplines. For now, Robin should be valued as a well-grounded demonstration of semi-autonomous hypothesis-to-analysis orchestration, not as a general proof that distributed agents can autonomously discover therapies.

DECLARATIONS

Authors' contributions

Conceptualization, writing and editing, literature analysis: Xing, X.; Zhang, JY.
Writing, literature analysis: Wang, J.; Celaje, J.

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tool ChatGPT (version 5.5 Pro, released 2026-04-23) was used solely for language editing. The tool did not influence the study design, data collection, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Financial support and sponsorship

None.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

REFERENCES

1. Ghareeb, A. E.; Chang, B.; Mitchener, L.; et al. A multi-agent system for automating scientific discovery. *Nature* **2026**, *655*, 497-505. DOI PubMed PMC
2. Buehler, M. J. Generative retrieval-augmented ontologic graph and multiagent strategies for interpretive large language model-based materials design. *ACS. Eng. Au.* **2024**, *4*, 241-77. DOI PubMed PMC
3. Boiko, D. A.; MacKnight, R.; Kline, B.; Gomes, G. Autonomous chemical research with large language models. *Nature* **2023**, *624*, 570-8. DOI PubMed PMC
4. Lu, C.; Lu, C.; Lange, R. T.; et al. Towards end-to-end automation of AI research. *Nature* **2026**, *651*, 914-9. DOI PubMed PMC
5. Wang, F. Y.; Marom, L.; Pal, S.; et al. Autonomous agents coordinating distributed discovery through emergent artifact exchange. *arXiv* **2026**, arXiv:2603.14312. Available online: https://doi.org/10.48550/arXiv.2603.14312 (accessed 28 September 2026).
6. Skarlinski, M. D.; Cox, S.; Laurent, J. M.; et al. Language agents achieve superhuman synthesis of scientific knowledge. *arXiv* **2024**, arXiv:2409.13740. Available online: https://doi.org/10.48550/arXiv.2409.13740 (accessed 28 September 2026).
7. Mao, Y.; Finnemann, S. C. Acute RhoA/Rho kinase inhibition is sufficient to restore phagocytic capacity to retinal pigment epithelium lacking the engulfment receptor MerTK. *Cells*. **2021**, *10*, 1927. DOI PubMed PMC
8. Storti, F.; Klee, K.; Todorova, V.; et al. Impaired ABCA1/ABCG1-mediated lipid efflux in the mouse retinal pigment epithelium (RPE) leads to retinal degeneration. *ELife* **2019**, *8*, e45100. DOI PubMed PMC

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.

© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.