On the Horizon

Future Molecular Science

Organized Open-weight AI Agents for Chemical Research: From Model Relay to Digital Research Organizations

Yunkai Hu[1#](胡运恺), Zhiyan Liang[1#](梁志燕), Jiazhen Cao[1](曹嘉真), Yichi Zha[1](查一驰), Daqing Jia[2,3*](贾大庆), Yue Jiang[1,2*](姜越), Mingyang Xing[1,2*](邢明阳)

[1] Key Laboratory of Green Chemical Engineering and Industrial Catalysis, School of Chemistry & Molecular Engineering, East China University of Science and Technology, Shanghai 200237, China
[2] National Engineering Research Center of Industrial Wastewater Detoxication and Resource Recovery, East China University of Science and Technology, Shanghai 200237, China
[3] School of Resources and Environmental Engineering, East China University of Science and Technology, 130 Meilong Road, Shanghai 200237, China

Abstract

Artificial intelligence is rapidly entering chemical research, yet most current applications remain fragmented: individual models assist with isolated tasks but lack persistent memory, coordinated workflows, and a shared representation of scientific evidence. We argue that further progress will depend less on making individual models more capable than on organizing models, tools, and evidence into coherent research systems. We propose an organized open-weight agent framework in which a research-planning model decomposes scientific questions and coordinates specialized agents for literature analysis, chemical reasoning, code development, structure construction, computational execution, and critical validation. Tool-call interfaces connect these agents to domain-specific chemistry software, while a shared evidence blackboard records hypotheses, supporting and contradictory evidence, computational parameters, failed attempts, and code versions, so that the research state persists and accumulates over time. Individual models thus become replaceable components rather than repositories of scientific context, and the objective shifts from generating answers quickly to selecting the next action that most effectively reduces scientific uncertainty. In this division of labor, models reason and coordinate, tools execute, evidence constrains and adjudicates claims, and human scientists define objectives and retain authority over critical decisions. We envision a transition from model relay to persistent digital research organizations, making AI-assisted chemical research more systematic, reproducible, and capable of addressing complex scientific questions over time.

1 Introduction

Artificial intelligence has long supported retrosynthetic planning,[1] robotic synthesis,[2,3] and property prediction,[4] and contemporary agents can use tools, coordinate modules, and retain files.[5-9] Yet projects still lose the rationale that connects hypotheses, contradictory observations, failures, and later decisions. The remaining challenge is the continuity and accountability of scientific reasoning across tasks, tools, and model versions.

This challenge becomes acute when mechanisms remain contested. In persulfate-based advanced oxidation, similar catalytic systems have been attributed to radical, singlet-oxygen, or surface-mediated pathways, prompting calls for stricter evidentiary standards.[10-13] Interfacial electron transfer and mass-transfer conditions can further alter reactive-oxygen-species generation and degradation performance.[14,15] Resolving such disputes requires that calculations, observations, and counterarguments remain linked within a single auditable research state.

We therefore argue that better models must be matched by better organization of models, tools, and evidence. Autonomous laboratories and large-scale materials discovery already couple prediction to synthesis and candidate assessment,[16,17] and the transition from language models to scientific agents has been reviewed elsewhere.[18] We define a digital research organization as a persistent, governed process in which hypotheses, evidence, tasks, and decisions remain identifiable independently of the models that perform them. Our contribution is a reference design that integrates evidence-quality assessment, conflict-preserving updates, uncertainty-guided task selection, and human authority (Figure 1). We do not claim to invent tool use, memory, multi-agent collaboration, or autonomous experimentation, nor do we report an implemented platform or measured performance gains.

Received: August 28, 2026; Revised: September 21, 2026;
Accepted: October 13, 2026; Published Online: October 20, 2026
*Corresponding authors: Daqing Jia (jiadaqing@ecust.edu.cn), Yue Jiang (yue0801@ecust.edu.cn), Mingyang Xing (mingyangxing@ecust.edu.cn)
#These authors contributed equally to this work.

Check for updates

DOI: 10.68642/ftms.202600003

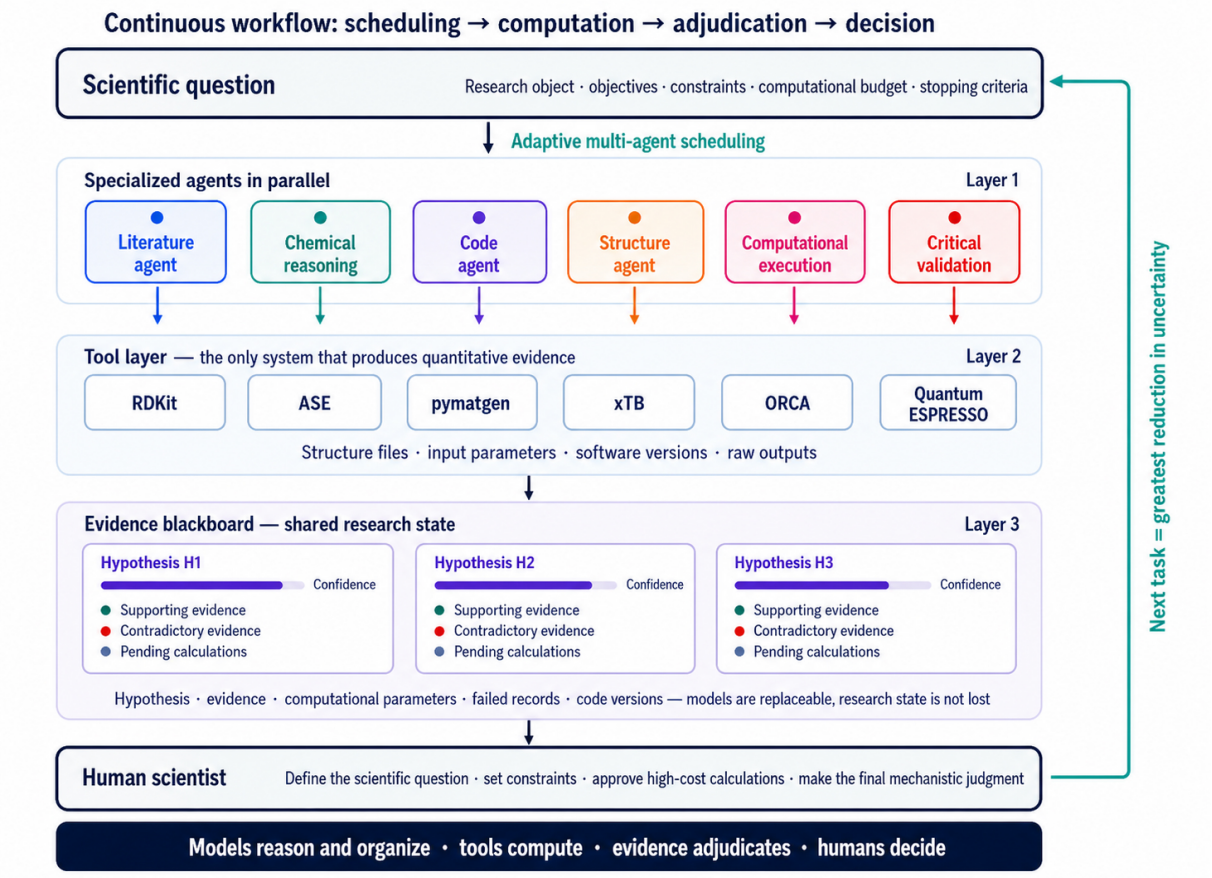


Figure 1 **Continuous workflow of the proposed organization.** The research-planning model decomposes a scientific question into tasks; the tool layer comprises computational software, database access, and approved experimental interfaces. Confidence bars denote reviewed evidentiary states rather than model-generated probabilities.

2 From Model Relay to Digital Research Organizations

A language model proposes explanations and tool calls; an agent couples such a model to an execution environment through defined interfaces.[19] Persistent workspaces can retain files and run histories, but they do not by themselves determine how scientific claims should evolve. In the proposed organization, the durable unit is not the latest conversational answer but a hypothesis linked to observations, methods, alternatives, and decisions.

A planning model schedules independent branches while respecting dependencies, and a version-controlled workspace retains scripts, inputs, outputs, and failures. Every decision records which evidence changed the research state and why, and these records—including unresolved disagreements—must survive model replacement. Executable calculations, supported interpretations, and approved conclusions remain distinct categories of outcome.

Digital materials ecosystems already connect databases, predictive models, simulations, experiments, and feedback.[20,21] The proposed organization operates within such ecosystems as a project-level coordination and evidence-governance layer: it does not replace laboratory infrastructure, databases, active learning, or other task-selection algorithms, but specifies the evidence, permissions, and review required to accept an action and retain its consequences.

Related perspectives emphasize mechanistic reasoning, physicochemical constraints, knowledge accumulation, and the evaluation of executable workflows.[22,23] Building on these aims, we specify how evidence is admitted, how conflicting interpretations are preserved, and how subsequent actions are justified. Table 1 compares reported emphases with the proposed requirements; unreported capabilities should not be read as absent, and no performance ranking is implied.

3 Open-weight Models: Advantages, Trade-offs, and Governance

Open-weight models are a deployment option, not a prerequisite. Accessible checkpoints support local use, inspection, adaptation, replacement, and archiving,[24] subject to licensing terms and available resources. These benefits must be weighed against task capability, hardware and maintenance costs, security, and the need for revalidation after model changes. Closed models can participate in hybrid workflows where policy permits, and open weights alone do not guarantee safety, permissive licensing, or reproducibility.

Model accessibility, workflow reproducibility, and computational reproducibility are distinct properties. Replaying a workflow requires recorded prompts, task dependencies, tool calls, code, inputs, outputs, environments, and data snapshots; repeat-

DOI: 10.68642/ftms.202600003

Table 1 **Positioning of the proposed organization relative to representative scientific-agent and discovery work.**

Representative work	Reported emphasis	Relationship to this proposal
Chemistry tool agents[5] and autonomous chemical research[6]	Tool-augmented reasoning and executable scientific actions.	Established foundations; our focus is on explicit, persistent records of hypotheses, evidence, and decisions.
Modular materials agents[9]	Specialized components for computational materials tasks.	Role modularity is inherited; evidence admission and conflict updates are additional design requirements here.
Autonomous laboratory[16]	Closed-loop synthesis and characterization of materials.	Experimental infrastructure can supply evidence and execute approved tasks.
Digital materials ecosystem[20,21]	Integration of databases, AI, simulations, experiments, and feedback.	Broader enabling ecosystem; the proposed organization manages project-level research state.
Reasoning loops;[22] executable-workflow review[23]	Mechanistic knowledge, physical constraints, and rigorous workflow evaluation.	Substantial conceptual overlap; our contribution specifies evidence and decision records.
This Perspective	Proposed rules for evidence quality, conflict handling, task choice, and human review.	Conceptual reference architecture; empirical benefits remain to be evaluated.

ing a calculation additionally requires agreement within declared numerical tolerances. Model and tokenizer versions, inference settings, quantization, hardware, and libraries can all matter. Exact textual reproduction is neither always achievable nor synonymous with scientific robustness, and FAIR provenance organizes these records without guaranteeing reproducibility.[25]

Governance applies equally to open and closed models. Institutions should define role-based permissions, approved tools, data boundaries, budgets, and accountable human owners. High-risk laboratory actions, safety-related changes, and commitments beyond approved resources require explicit authorization enforced at the tool level. Local deployment should enable controlled, auditable research rather than bypass safeguards, and hybrid workflows must retain accessible artifacts while disclosing non-replayable proprietary dependencies (SI Section S2).

4 Reference Architecture and Agent Harness

The research-planning model converts objectives, constraints, hypotheses, available tools, budgets, and stopping criteria into a task queue. Six reference roles—literature analysis, chemical reasoning, coding, structure construction, computational execution, and critical validation—cover the main workflow (Table 2); these are functions, not a universal taxonomy or a requirement for six separate models. The tool layer may connect RDKit, ASE,[26] pymatgen,[27] xTB,[28] ORCA,[29] Quantum ESPRESSO,[30] the Materials Project,[31] and machine-learning interatomic potentials.[32] Quantitative claims are admissible only when supported by documented calculations, measurements, or curated data.

The agent harness couples planning to execution: it validates tool inputs and outputs, enforces dependencies, permissions, and budgets, and manages bounded retries, checkpoints, and review gates. The harness controls execution; the Evidence Blackboard preserves scientific state. A successful process exit does not by itself admit a result as evidence—convergence, units, applicability, and provenance must first pass review. Failures remain visible, and repeated validation or resource failures trigger a pause or escalation.

The roles form an evidentiary chain: literature and reasoning agents convert traceable observations into testable alternatives; structure and code agents turn these alternatives into reproducible tasks; execution agents generate quantitative outputs; and validation agents test quality and counterevidence. Literature extraction should span text, tables, figures, and supporting information while preserving location, identity, conditions, units, and uncertainty, and source observations, author interpretations, and agent inferences must remain distinct. DIVE-to-design illustrates how figure-centered literature can be converted into structured knowledge.[33] Evidence links take one of four relations—supports, contradicts, non-discriminating, or not comparable (SI Section S3).

Execution agents may invoke screening models, xTB, DFT, or molecular dynamics while archiving inputs, outputs, and quality checks; validation agents probe for omitted pathways, unit errors, nonconvergence, and limited applicability. The nanocluster workflow of Cai et al., which combines neural-network potentials with unsupervised analysis,[34] illustrates a callable computational capability rather than an organizational taxonomy. DFT is not ground truth, and convergence alone does not establish relevance to the experimental system.

Domain-specific interfaces extend the architecture beyond computation: experimental-design and data-analysis roles can propose discriminating controls and interpret spectra, and approved instrumentation interfaces can connect protocols to characterization or robotic synthesis. Records should link sample identity, preparation, calibration, acquisition settings, raw signals, and processing steps. Physical interlocks and institutional procedures remain authoritative, and projects may keep experiments fully human-operated (SI Section S1).

DOI: 10.68642/ftms.202600003

Table 2 **Reference specialized-agent roles and primary responsibilities; roles may be combined or extended for a particular domain.**

Specialized agent	Primary responsibilities
Literature agent	Identifies primary articles, databases, and supporting information for the target system; extracts experimental conditions, structures, characterization data, performance metrics, and mechanistic evidence; retains traceable sources, locations, and applicability conditions for critical conclusions; and flags conflicting reports as evidence requiring verification. Converts text, figures, and tables into normalized records, distinguishing source observations from author and agent interpretations.
Chemical reasoning agent	Proposes distinguishable reaction pathways, key intermediates, and rate-controlling steps on the basis of literature evidence and known constraints; checks charge, mass, spin, coordination, and chemical plausibility; and specifies the calculations or experiments required for each hypothesis.
Code agent	Translates research questions into reproducible data-processing and computational workflows; writes scripts for structure generation, job submission, result parsing, testing, and visualization; records dependencies, versions, and run logs; and promptly exposes failures and retry conditions.
Structure agent	Generates molecules, crystals, surfaces, defects, coordination sites, and adsorption configurations from the literature, species states, and surface or crystal constraints; enumerates key initial configurations; and annotates charge, spin, coverage, and boundary conditions for subsequent screening and calculation.
Computational execution agent	Invokes machine-learning interatomic potentials, xTB, DFT, or molecular dynamics according to task priority to perform geometry optimization and calculations of energies, reaction pathways, and kinetics; archives input files, parameters, software versions, raw outputs, and quality-control results; and produces reviewable quantitative evidence.
Critical validation agent	Actively seeks counterexamples, omitted pathways, and inconsistent evidence; rechecks parameters, units, convergence, model applicability, and reasoning chains; grades evidence strength and uncertainty; and recommends additional calculations, experiments, or appropriately qualified conclusions.

5 Evidence Blackboard: From Records to Reviewed Claims

The Evidence Blackboard is a versioned research-state layer, not a transcript or an agent vote. Persistent hypothesis records define scope, alternatives, status, and linked evidence; evidence records separate observations from interpretations; and decision records explain why a state or task changed. Sources, methods, conditions, uncertainty, validation, and dependencies all remain traceable, and corrections supersede earlier interpretations without deleting the original evidence (SI Section S4).

Evidence is assessed along five dimensions: provenance, methodological adequacy, condition applicability, discriminatory power, and independence. No universal hierarchy of experiment > DFT > surrogate > literature is appropriate, because the literature may report any underlying method. A controlled measurement can directly discriminate among mechanisms, whereas a nonspecific probe cannot; a DFT result is informative only within its structural and methodological assumptions. This distinction is especially important for probe-based assignments in advanced oxidation and for condition-dependent pathway changes.[35-37]

When records disagree, the organization first checks material identity, conditions, observables, units, and time scales, and then examines method validity and shared dependencies. Repeated citations of the same experiment, or predictions from the same model family, do not constitute independent confirmation. Credible contradictions remain linked and can place a hypothesis in a contested state; failed or inapplicable calculations are retained as failures, not as evidence of chemical impossibility (SI Section S5).

Confidence is a reviewed evidentiary state, not model self-confidence. We suggest five states: untested, provisionally supported, strongly supported within scope, contested, and disfavored within scope. Each transition records the evidence, rationale, reviewer, and unresolved alternatives. Bayesian updating is appropriate only when priors, likelihoods, and dependence assumptions can be justified; otherwise, ordinal states are more defensible. Humans retain authority over final mechanistic conclusions and consequential changes.

6 Selecting Actions That Reduce Scientific Uncertainty

Scientific uncertainty takes several forms: competing explanations, imprecise decision-relevant parameters, and uncertain model or measurement applicability. An informative action is one expected to distinguish among alternatives, narrow a relevant interval, or change a research decision. Information gain, decision value, expected reliability, and cost should be weighed only after permission, safety, and feasibility checks. Generating more text, or expressing greater model confidence, does not reduce scientific uncertainty.

Conceptually, the system selects, from among permitted actions, the one offering the greatest expected reduction in a declared uncertainty measure after an explicit cost penalty. Entropy is suitable for calibrated, mutually exclusive hypotheses; coexisting pathways require contribution-based or joint representations. When calibrated probabilities are unavailable, a documented qualitative ranking of discrimination, reliability, decision relevance, and cost is preferable. This is a design princi-

DOI: 10.68642/ftms.202600003

ple rather than an implemented optimization algorithm; assumptions and stopping rules are detailed in SI Section S6.

Figure 2 contrasts this organized system with isolated agent use, and SI Section S7 details a hypothetical inquiry into PMS activation. In that example, reports may support freely diffusing radicals, singlet oxygen, or surface-mediated oxidation under incompletely matched conditions, and the pathways may coexist. The organization separates observations from interpretations, checks condition matching and probe limitations, and preserves unresolved alternatives; it then weighs provenance checks, discriminating controls, orthogonal measurements, and targeted calculations, selecting the feasible action most likely to resolve the current decision. Outcomes may support a scoped explanation, reveal mixed pathways, or remain inconclusive; no new data, barriers, or benchmark results are claimed.

7 Evaluating the Organization

Evaluation must separate execution success from scientific reliability. We propose the following measures: success against predefined scientific criteria; error-detection recall, precision, and false-positive rate; provenance completeness; rerun agreement within declared tolerances; computational, experimental, and human cost; and planned versus unplanned human intervention. Probability calibration should be tested whenever numerical confidence is reported. SI Section S8 defines denominators and failure cases; these are benchmark proposals, not results.

Future studies should compare a documented model-relay baseline against the full organization under matched models, tools, evidence, tasks, and budgets, with ablations that remove evidence links, conflict review, or uncertainty-guided scheduling. Test suites should include duplicated or condition-mismatched evidence, nonconvergent calculations, omitted alternatives, interruptions, and model replacement, and outcomes should be assessed through expert review, repeated runs, held-out tasks, and complete accounting of human effort.

8 Conclusions and Outlook

A digital research organization preserves why claims were supported, challenged, revised, or deferred. Models serve as replaceable reasoning components, tools generate observations, the Evidence Blackboard preserves the history of evidence and interpretation, and the harness enforces execution and review. The proposal remains conceptual and requires domain-specific calibration and empirical testing; we close with five priorities.

(i) Standardized research-state objects. Machine-readable hypotheses, evidence, tasks, and calculations need persistent identifiers so that research state can move across models and groups.

(ii) Organizational benchmarks. Evaluations should compare accepted outcomes, error detection, provenance, rerun agreement, cost, and coordination overhead under matched resources—and should allow for the possibility that simple tasks gain nothing.

(iii) Physics-grounded reasoning. Conservation laws, admissible chemical states, and structural constraints should filter proposals, while thermodynamics, kinetics, and operating conditions constrain interpretation. Hard invariants must be distinguished from approximate models and heuristics, with explicit treatment of surfaces, defects, external driving, and open-system exchange. Independent calculations and experiments remain indispensable (SI Section S9).

(iv) Retention of negative results. Failed structures, nonconvergent calculations, and disfavored mechanisms should remain accessible, because they constrain future decisions.

(v) Human-review gates. Humans define objectives and permissions, approve high-risk operations and budget exceptions, and retain final mechanistic authority. The audit trail records the approver, scope, time, and supporting evidence; neither an agent nor the harness can grant itself new authority.

An auditable, semi-automated workflow is a practical starting point. Its value should be judged by whether it yields better-supported, reproducible decisions over time, while preserving uncertainty, negative results, and accountability.

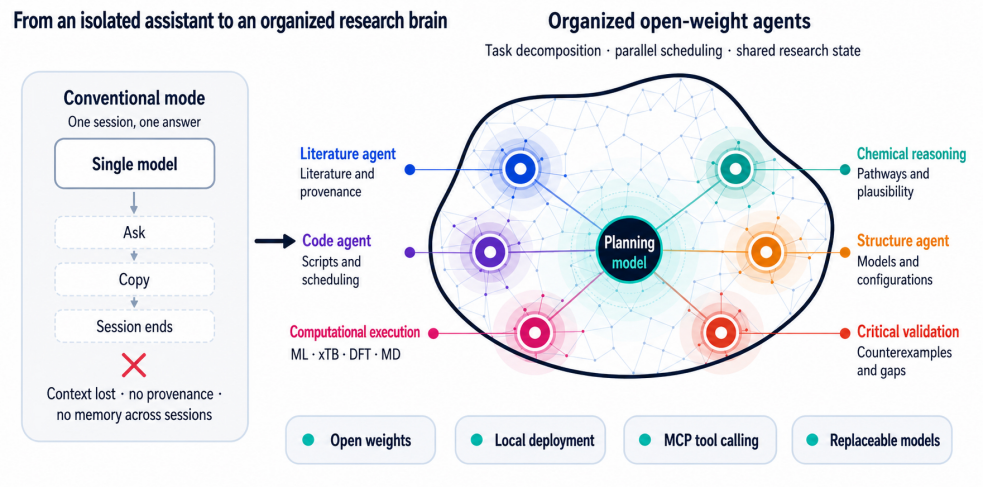


Figure 2 **From conventional single-model assistance to an organized agent system.**

Author Contributions

Yunkai Hu: Conceptualization, Writing – Original Draft. Zhiyan Liang: Conceptualization, Writing – Original Draft. Jiazhen Cao: Investigation, Writing – Review & Editing. Yichi Zha: Investigation, Writing – Review & Editing. Daqing Jia: Supervision, Writing – Review & Editing, Funding Acquisition. Yue Jiang: Supervision, Writing – Review & Editing, Funding Acquisition. Mingyang Xing: Conceptualization, Supervision, Writing – Review & Editing, Funding Acquisition.

Supporting Information

Supplementary data associated with this article can be found, in the online version, at DOI: 10.68642/ftms.202600003.

Data Availability Statement

No new datasets were generated or analyzed in this study.

Competing Interests

The authors declare no competing financial or non-financial interests.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 22325602, 22521201, 22176060, and 22406142), the National Key R&D Program of China (Grant No. SQ2024YFA1211001), the Program of Shanghai Academic/Technology Research Leader (Grant Nos. 23XD1421000 and 25ZR1402099), the Chenguang Program of Shanghai Education Development Foundation and Shanghai Municipal Education Commission (Grant No. 25CGA25), the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (Grant No. JYB2025XDXM404), the Key Laboratory of Functional Inorganic Material Chemistry (Heilongjiang University), Ministry of Education (Grant No. KIM-0-2025017), the State Key Laboratory of Water Pollution Control and Green Resource Recycling Foundation (Grant No. PCRRF250014), and the Shanghai Pujiang Talent Program (Grant No. 25PJD025).

This is an open access article under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0, https://creativecommons.org/licenses/by/4.0/).
© The Author(s) 2026.

References and Notes

1. Segler, M. H. S., Preuss, M. & Waller, M. P. Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI. *Nature* **2018**, 555, 604–610. DOI: 10.1038/nature25978
2. Coley, C. W., Thomas, D. A., Lummiss, J. A. M., Jaworski, J. N., Breen, C. P., Schultz, V., Hart, T., Fishman, J. S., Rogers, L., Gao, H., *et al*. A Robotic Platform for Flow Synthesis of Organic Compounds Informed by AI Planning. *Science* **2019**, 365, eaax1566. DOI: 10.1126/science.aax1566
3. Burger, B., Maffettone, P. M., Gusev, V. V., Aitchison, C. M., Bai, Y., Wang, X., Li, X., Alston, B. M., Li, B., Clowes, R., *et al*. A Mobile Robotic Chemist. *Nature* **2020**, 583, 237–241. DOI: 10.1038/s41586-020-2442-2
4. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, 559, 547–555. DOI: 10.1038/s41586-018-0337-2
5. Bran, A. M., Cox, S., Schilter, O., Baldassari, C., White, A. D. & Schwaller, P. Augmenting Large Language Models with Chemistry Tools. *Nat. Mach. Intell.* **2024**, 6, 525–535. DOI: 10.1038/s42256-024-00832-8
6. Boiko, D. A., MacKnight, R., Kline, B. & Gomes, G. Autonomous Chemical Research with Large Language Models. *Nature* **2023**, 624, 570–578. DOI: 10.1038/s41586-023-06792-0
7. Jablonka, K. M., Schwaller, P., Ortega-Guerrero, A. & Smit, B. Leveraging Large Language Models for Predictive Chemistry. *Nat. Mach. Intell.* **2024**, 6, 161–169. DOI: 10.1038/s42256-023-00788-1
8. Zheng, Z., Zhang, O., Borgs, C., Chayes, J. T. & Yaghi, O. M. ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis. *J. Am. Chem. Soc.* **2023**, 145, 18048–18062. DOI: 10.1021/jacs.3c05819
9. Chaudhari, A., Ock, J. & Barati Farimani, A. Modular Large Language Model Agents for Multi-task Computational Materials Science. *Commun. Mater.* **2026**, 7, 131. DOI: 10.1038/s43246-025-00994-x
10. Zhou, H., Ren, W., Wang, S. & Duan, X. Best Practices for Advanced Oxidation Processes. *Nat. Sustain.* **2025**, 8, 1428–1431. DOI: 10.1038/s41893-025-01655-0
11. Zhang, S., Zheng, H. & Tratnyek, P. G. Advanced Redo$_x$ Processes for Sustainable Water Treatment. *Nat. Water* **2023**, 1, 666–681. DOI: 10.1038/s44221-023-00098-1
12. Lee, J., von Gunten, U. & Kim, J. -H. Persulfate-Based Advanced Oxidation: Critical Assessment of Opportunities and Roadblocks. *Environ. Sci. Technol.* **2020**, 54, 3064–3081. DOI: 10.1021/acs.est.9b07082
13. Yan, Y., Wei, Z., Duan, X., Long, M., Spinney, R., Dionysiou, D. D., Xiao, R. & Alvarez, P. J. J. Merits and Limitations of Radical vs. Nonradical Pathways in Persulfate-Based Advanced Oxidation Processes. *Environ. Sci. Technol.* **2023**, 57, 12153–12179. DOI: 10.1021/acs.est.3c05153
14. Liang, L., Cao, J., Chen, Z., Liang, Z., Jiang, Y. & Xing, M. Harnessing Fe–Mo Atomic Interfaces for Boosted Electron Transfer and ROS Generation in Sustainable Pollutant Degradation. *Environ. Sci. Technol.* **2025**, 59, 22914–22926. DOI: 10.1021/acs.est.5c06307
15. Fu, J., Xiao, S., Cao, J., Liang, Z., Chen, J., Jiang, Y. & Xing, M. Mass Transfer -Enhanced Photothermal Membranes with Synergistic Light Utilization for High -Turbidity Wastewater Purification. *Angew. Chem. Int. Ed.* **2025**, 64, e202421800. DOI: 10.1002/anie.202421800
16. Szymanski, N. J., Rendy, B., Fei, Y., Kumar, R. E., He, T., Milsted, D., McDermott, M. J., Gallant, M., Cubuk, E. D., Merchant, A., *et al*. An Autonomous Laboratory for the Accelerated Synthesis of Inorganic Materials. *Nature* **2023**, 624, 86–91. DOI: 10.1038/s41586-023-06734-w
17. Merchant, A., Batzner, S., Schoenholz, S. S., Aykol, M., Cheon, G. & Cubuk, E. D. Scaling Deep Learning for Materials Discovery. *Nature* **2023**, 624, 80–85. DOI: 10.1038/s41586-023-06735-9
18. Yao, T., Huang, J., Yan, Y., Yang, Y., Wang, Z., Shao, X., Gao, Z. & Yang, W. From Large Language Models to AI Agents in Energy Materials Research: Enabling Discovery, Design, and Automation. *AI Agent* **2025**, 1, 9. DOI: 10.20517/aiagent.2025.03
19. Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., Cancedda, N. & Scialom, T. Toolformer: Language Models Can Teach Themselves to Use Tools. *Adv. Neural Inf. Process. Syst.* **2023**, 36, 68539–68551. DOI: 10.48550/arXiv.2302.04761
20. Zhang, D., Jia, X., Wang, Y., Liu, H., Wang, Q., Jang, S. -H., Shah, D., Ye, S., Tran, H. B. & Li, H. Digital Materials Ecosystem: from Databases to AI Agents for Autonomous Discovery. *Chem. Sci.* **2026**, 17, 5782–5804. DOI: 10.1039/d5sc09229a
21. Zhang, D., Chen, Y., Liu, C., Liu, Y., Xin, H., Peng, J., Ou, P. & Li, H. Accelerating Catalyst Materials Discovery With Large Artificial Intelligence Models. *Angew. Chem. Int. Ed.* **2026**, 65, e26150. DOI: 10.1002/anie.202526150
22. Tan, Z., Qian, L., Wang, Y. & Ran, N. From Execution Loop to Reasoning Loop in Autonomous Materials Research. *AI Agent* **2026**, 2, 13. DOI: 10.20517/aiagent.2026.13
23. Hu, C. & Wang, Z. Computational Materials Agents: from Task Demonstrations to Executable Scientific Workflows. *AI Agent* **2026**, 2, 21. DOI: 10.20517/aiagent.2026.38
24. Guo, D., Yang, D., Zhang, H., Song, J., Wang, P., Zhu, Q., Xu, R., Zhang, R., Ma, S., Bi, X., *et al*. DeepSeek-R1 Incentivizes Reasoning in LLMs through Reinforcement Learning. *Nature* **2025**, 645, 633–638. DOI: 10.1038/s41586-025-09422-z
25. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. -W., da Silva Santos, L. B., Bourne, P. E., *et al*. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, 3, 160018. DOI: 10.1038/sdata.2016.18
26. Hjorth Larsen, A., Jørgen Mortensen, J., Blomqvist, J., Castelli, I. E., Christensen, R., Dułak, M., Friis, J., Groves, M. N., Hammer, B., Hargus, C., *et al*. The Atomic Simulation Environment—a Python Library for

Working with Atoms. *J. Phys. : Condens. Matter* **2017**, 29, 273002. DOI: 10.1088/1361-648x/aa680e
27. Ong, S. P., Richards, W. D., Jain, A., Hautier, G., Kocher, M., Cholia, S., Gunter, D., Chevrier, V. L., Persson, K. A. & Ceder, G. Python Materials Genomics (Pymatgen): A Robust, Open-source Python Library for Materials Analysis. *Comput. Mater. Sci.* **2013**, 68, 314–319. DOI: 10.1016/j.commatsci.2012.10.028
28. Bannwarth, C., Ehlert, S. & Grimme, S. $GFN_{2\text{-}x}TB$—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, 15, 1652–1671. DOI: 10.1021/acs.jctc.8b01176
29. Neese, F., Wennmohs, F., Becker, U. & Riplinger, C. The ORCA Quantum Chemistry Program Package. *J. Chem. Phys.* **2020**, 152, 224108. DOI: 10.1063/5.0004608
30. Giannozzi, P., Baroni, S., Bonini, N., Calandra, M., Car, R., Cavazzoni, C., Ceresoli, D., Chiarotti, G. L., Cococcioni, M., Dabo, I., *et al*. QUANTUM ESPRESSO: a Modular and Open-source Software Project for Quantum Simulations of Materials. *J. Phys. : Condens. Matter* **2009**, 21, 395502. DOI: 10.1088/0953-8984/21/39/395502
31. Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., *et al*. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, 1, 011002. DOI: 10.1063/1.4812323
32. Deng, B., Zhong, P., Jun, K., Riebesell, J., Han, K., Bartel, C. J. & Ceder, G. CHGNet as a Pretrained Universal Neural Network Potential for Charge-informed Atomistic Modelling. *Nat. Mach. Intell.* **2023**, 5, 1031–1041. DOI: 10.1038/s42256-023-00716-3
33. Hong, Y. & Mao, X. DIVE-to-design: How a Multi-agent Workflow Converts Figure-centric Literature into an AI-native Hydrogen Storage Discovery Engine. *AI Agent* **2026**, 2, 2. DOI: 10.20517/aiagent.2026.04
34. Cai, H., Xu, A., Rajarathnam, G., Cao, A. & Yan, J. Exploring Structures of Nanoclusters Combining High-dimensional Neural Network Potentials with Unsupervised Machine Learning Algorithms. *AI Agent* **2025**, 1, 6. DOI: 10.20517/aiagent.2025.09
35. Li, H., Zhang, X., Yang, S., Sun, Y. & Qian, J. Discerning the Relevance of Singlet Oxygen in Pollutant Degradation in Peroxymonosulfate Activation Processes. *Environ. Sci. Technol.* **2024**, 58, 14005–14012. DOI: 10.1021/acs.est.4c02809
36. Liu, W., Lu, Y., Dong, Y., Jin, Q. & Lin, H. A Critical Review on Reliability of Quenching Experiment in Advanced Oxidation Processes. *Chem. Eng. J.* **2023**, 466, 143161. DOI: 10.1016/j.cej.2023.143161
37. Meng, Y., Liu, Y. -Q., Wang, C., Si, Y., Wang, Y. -J., Xia, W. -Q., Liu, T., Cao, X., Guo, Z. -Y., Chen, J. -J., *et al*. Nanoconfinement Steers Nonradical Pathway Transition in Single Atom Fenton-like Catalysis for Improving Oxidant Utilization. *Nat. Commun.* **2024**, 15, 5314. DOI: 10.1038/s41467-024-49605-2