

Rapid #: -27601484

CROSS REF ID: 22865708290002422

LENDER: PAU (University of Pennsylvania) :: Van Pelt

BORROWER: EQF (Bangor University) :: Main Library

TYPE: Article CC:CCG

JOURNAL TITLE: Matter

USER JOURNAL TITLE: Matter

ARTICLE TITLE: Knowledge-driven autonomous materials research via collaborative multi-agent and robotic system

ARTICLE AUTHOR: Shi, Tongyu

VOLUME: 9

ISSUE: 2

MONTH:

YEAR: 2026

PAGES: 102577

ISSN: 2590-2393

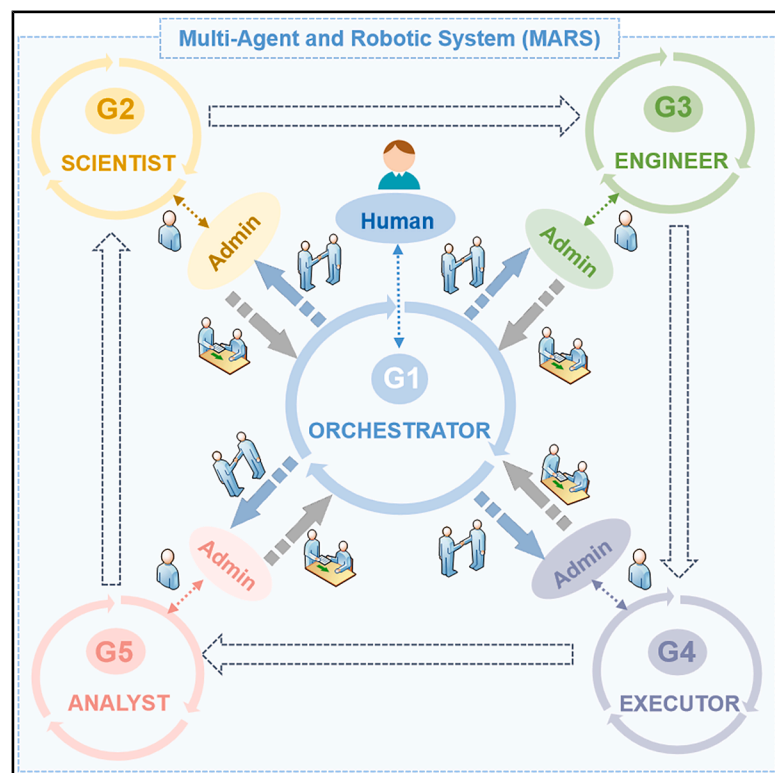
OCLC #:

Processed by RapidX: 10/2/2026 9:50:59 AM

This material may be protected by copyright law (Title 17 U.S. Code)

Knowledge-driven autonomous materials research via collaborative multi-agent and robotic system

Graphical abstract



Authors

Tongyu Shi, Yutang Li, Zhanlong Wang, ..., Paul K. Chu, Wenhua Zhou, Xue-Feng Yu

Correspondence

gl.jiang@siat.ac.cn (G.J.),
xf.yu@siat.ac.cn (X.-F.Y.)

In brief

The collaborative multi-agent and robot system (MARS), a hierarchical knowledge-driven framework coordinating 19 LLM agents with 16 domain-specific tools, is developed to accelerate materials research. By coordinating specialized agents with domain-specific tools, MARS autonomously executes full-cycle materials development, from design to synthesis and optimization. The system demonstrated a 60-fold acceleration in developing a water-stable perovskite nanocomposite through cross-domain knowledge integration. This paradigm shift enables researchers to focus on creative problem-solving while AI handles repetitive tasks, fundamentally changing materials innovation approaches.

Highlights

- Hierarchical multi-agent system with specialized tools mimics human research teams
- LLM-robot integration drives end-to-end autonomous materials research
- Hybrid retrieval enables cross-domain knowledge integration for professional designs
- Knowledge-based reasoning across domains leads to materials design



Benchmark

First qualification/assessment of material properties and/or performance



Shi et al., 2026, Matter 9, 102577
February 4, 2026 © 2025 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.
<https://doi.org/10.1016/j.matt.2025.102577>

Article

Knowledge-driven autonomous materials research via collaborative multi-agent and robotic system

Tongyu Shi,^{1,7} Yutang Li,^{1,7} Zhanlong Wang,^{1,7} Wenhe Xu,¹ Guolai Jiang,^{1,*} Dawei Dai,³ Jie Zhou,^{1,3} Hao Huang,¹ Rui He,^{1,2} Seeram Ramakrishna,⁴ Paul K. Chu,⁵ Wenhua Zhou,^{1,6} and Xue-Feng Yu^{1,2,6,8,*}

¹Materials Artificial Intelligence Center, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

²Science and Technology Institute of Advanced Technology, Wuhan 430050, China

³College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

⁴Department of Mechanical Engineering, National University of Singapore, Singapore, Republic of Singapore

⁵Department of Physics, Department of Materials Science and Engineering, and Department of Biomedical Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China

⁶Key Laboratory of Biomedical Imaging Science and System, Chinese Academy of Sciences, Shenzhen 518055, China

⁷These authors contributed equally

⁸Lead contact

*Correspondence: gl.jiang@siat.ac.cn (G.J.), xf.yu@siat.ac.cn (X.-F.Y.)

<https://doi.org/10.1016/j.matt.2025.102577>

PROGRESS AND POTENTIAL Large language models (LLMs) provide new possibilities for accelerating materials research, yet their application in complex materials science remains limited. Here, we developed the collaborative multi-agent and robot system (MARS), a knowledge-driven hierarchical architecture coordinating 19 LLM agents with 16 domain-specific tools organized into functional modules, achieving closed-loop autonomous materials discovery by integrating robotic experimentation. MARS offers accurate and professional guidance for materials development and relieves the hallucination inherent to current LLMs through a hybrid retrieval-augmented generation approach. As experimental validation, the system demonstrates its capabilities in perovskite nanocrystal development, optimizing synthesis protocols within 10 experimental iterations and designing a biomimetic “core-shell-corona” structure for water-stable perovskite composites within 3.5 h compared to conventional methods requiring 4–6 months. Through an interactive natural language interface, researchers can orchestrate experiments while tracking real-time agent coordination. Beyond efficiency gains, MARS represents a paradigm shift toward truly autonomous materials research where knowledge-driven AI can initiate scientific discovery. MARS can combine knowledge from different scientific fields to create solutions that human researchers might overlook, as demonstrated by its biomimetic design inspired by cell membranes. Looking forward, we envision MARS accelerating discoveries across critical areas including renewable energy materials, medical compounds, and sustainable technologies. This could help address urgent global challenges like climate change and healthcare needs by removing research bottlenecks. As the system evolves, it could democratize advanced materials research by reducing dependence on specialized infrastructure and expertise, allowing scientists worldwide to focus on creative problem-solving rather than repetitive tasks. MARS represents not just an improvement in research efficiency but also a fundamental shift in how we approach materials innovation.

SUMMARY

Large language models (LLMs) provide new possibilities for accelerating materials research, yet their application in complex materials science remains limited. Here, we developed the collaborative multi-agent and robot system (MARS), a knowledge-driven hierarchical architecture coordinating 19 LLM agents with 16 domain-specific tools for closed-loop autonomous materials discovery. MARS combines scientific knowledge with decision-making capabilities while mitigating hallucination through retrieval-augmented generation with a customized knowledge base. In experimental validation, the system



optimized perovskite nanocrystal synthesis within 10 iterations and designed a biomimetic “core-shell-corona” structure for water-stable perovskite composites in 3.5 h versus conventional methods requiring 4–6 months. This acceleration automates literature review and experimental planning, allowing researchers to focus on creative thinking while interacting through a natural language interface. This work establishes an integrated AI-driven framework for accelerating materials innovation and presents a paradigm for AI-enabled scientific discovery.

INTRODUCTION

Materials are the foundation of human understanding and transformation of the world, and innovations in this field are critical to societal progress and technological advancement.^{1,2} The discovery and application of new materials is typically a complex, lengthy, and costly process that requires the accumulation of multidisciplinary knowledge, extensive study of prior research, expertise in preparation techniques, time-intensive experimental iterations, and thorough validation in practical scenarios.^{3–6} To enhance the efficiency of materials innovation, scientists have made sustained and long-term efforts, such as the launch of the Materials Genome Initiative in 2011, which aims to accelerate materials discovery using high-throughput computational and experimental platforms.^{7–9} Despite these advances, the development of new materials typically still requires years, highlighting the urgent need for new methodologies to accelerate the pace of materials research.

Since OpenAI released ChatGPT, large language models (LLMs) have attracted widespread research interest, leading to substantial breakthroughs across various fields.^{10,11} LLMs can serve as invaluable research assistants and aid researchers in efficiently mining literature, quickly accessing critical data, and providing potential optimization strategies.^{12–15} Besides, LLMs have recently shown exceptional abilities in code generation, specialized tool integration, and strategic planning.^{16–23} These advances have enabled several pioneering studies that successfully couple LLM-powered agents with robotic experimentation platforms, achieving autonomous chemical synthesis,^{24,25} accelerated reaction optimization,²⁶ and intelligent synthesis route recommendation,²⁷ thereby demonstrating the transformative potential of AI-driven laboratory automation. These capabilities are particularly valuable for materials science, where the inherently multidisciplinary nature demands precise knowledge extraction and integration from an ever-expanding scientific literature.²⁸ This is particularly evident in nanomaterials research, a field characterized by an explosion of scientific literature over the past two decades, yet struggles with challenges of research redundancy and reproducibility.^{29,30} While promising, leveraging the full potential of LLMs for autonomous materials research requires addressing key challenges, including accurate comprehension of domain-specific knowledge, mitigation of hallucinations, robust control of automated synthesis, efficient management of complex tasks, and development of systematic workflows.

Herein, we develop a knowledge-driven multi-agent and robot system (MARS) for end-to-end autonomous materials discovery. MARS employs a collaborative network of 19 specialized LLM agents with access to 16 domain-specific tools to orchestrate complex material research tasks by integrating the robotic plat-

form. Critically, MARS leverages a comprehensive materials science knowledge base that guides its reasoning and mitigates hallucinations, providing professional guidance and designing multi-step synthesis processes compatible with automated robotic platforms. Subsequently, MARS can control multiple robotic platforms to autonomously execute iterative experiments (synthesis, analysis, and optimization) to achieve research objectives, demonstrated by successfully tuning green-emitting perovskite nanocrystals to target blue emission within 10 experimental iterations. These capabilities enable MARS to function as a human research team with superior domain-specific expertise compared to existing LLM-based systems. Furthermore, demonstrating its accelerated discovery potential, MARS autonomously designed and synthesized CsPbBr₃@LP/PEG composite with enhanced water stability and biocompatibility within 3.5 h, significantly accelerating materials development by 60-fold compared with the traditional method. By automating literature review, experimental planning, and protocol optimization, MARS allows human researchers to redirect their focus toward higher-level scientific insights, hypothesis generation, and creative problem-solving. Moreover, a user-friendly web interface facilitates intuitive interaction with MARS, enabling researchers to specify research objectives in natural language and monitor the automated workflow, creating a synergistic human-AI partnership that maximizes scientific productivity and accelerates the pace of materials innovation.

RESULTS

The architecture of MARS

This automated materials research framework employs a multi-agent system architecture that is analogous to a well-coordinated research laboratory. The MARS framework coordinates a network of 19 specialized agents, enabling complex multi-objective tasks in materials design, synthesis, and analysis through collaborative intelligence. These agents constitute a higher-level framework built on LLMs, leveraging their core reasoning and generative capabilities while incorporating task definition, dynamic utilization of 16 professional toolkits, and feedback loops to achieve complex objectives beyond the capacity of standalone LLMs. As shown in Figure 1A, the system is hierarchically structured into five synergistic functional groups with distinct responsibilities to emulate the real-world scientific research team, including the ORCHESTRATOR group (G1), SCIENTIST group (G2), ENGINEER group (G3), EXECUTOR group (G4), and ANALYST group (G5) (see Note S1).

The ORCHESTRATOR Group serves as the central coordinator, plans sub-tasks based on user requests, and selects the appropriate agent group to complete the sub-task, while

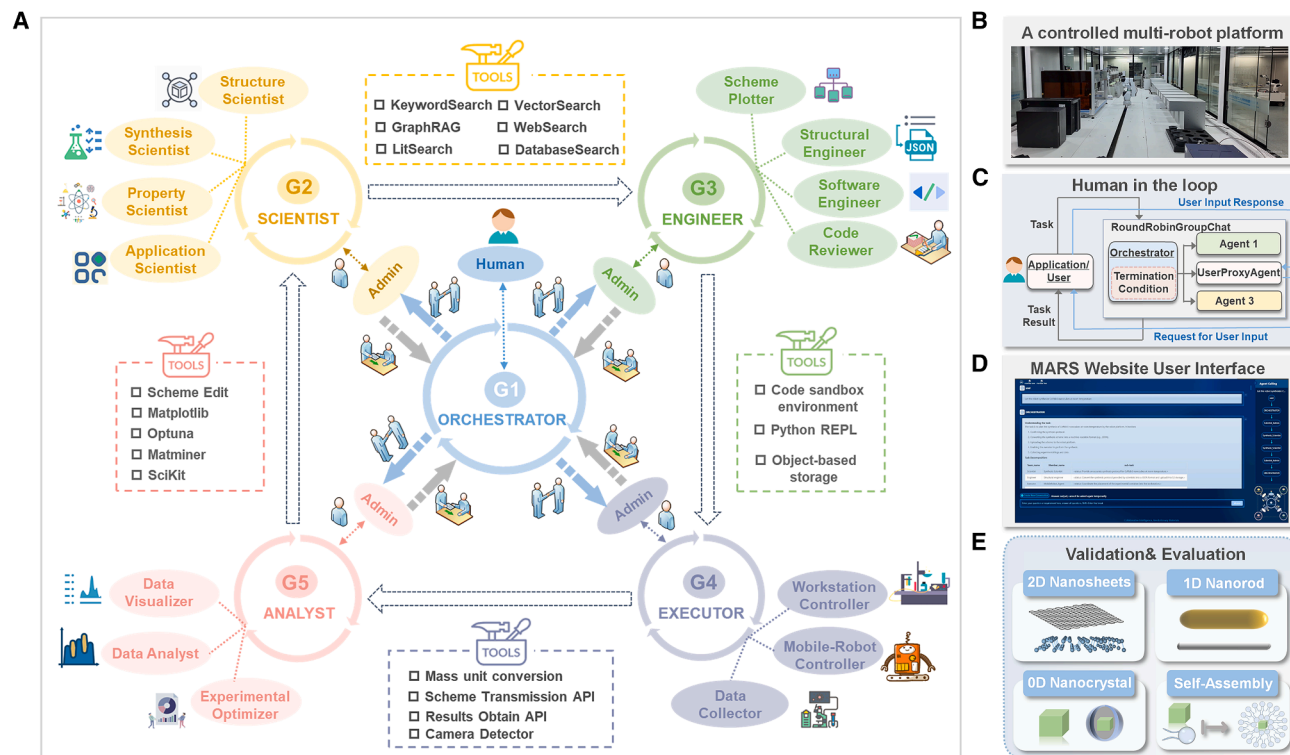


Figure 1. Architecture of MARS

(A) Overview of the MARS system. Circles represent 5 functional groups, and dashed boxes indicate callable toolkits.
(B) Physical setup of the multi-robot control platform.
(C) The mechanism of human-in-the-loop.
(D) Web-based user interface of MARS.
(E) Experimental validation and evaluation.

managing the message queue between agent groups. The SCIENTIST group, equipped with the hybrid retrieval-augmented generation (RAG) strategy and several customized information retrieval tools, focuses on materials knowledge retrieval and solution design, satisfying materials synthesis, structure, property, and application requirements. The ENGINEER group translates scientific proposals into executable protocols, bridging the gap between theoretical design and practical implementation. As presented in Figure 1B, the EXECUTOR group interacts with automated experimental platforms to carry out these protocols, controlling instrumentation, collecting data, and ensuring the reliable execution of designed experiments. Finally, the ANALYST Group analyzes experimental results and generates refinement strategies, driving iterative improvement and optimization based on experimental data. This modular design allows for specialized expertise and efficient task distribution among the agent groups.³¹

An effective, hierarchical message-passing mechanism is crucial to MARS, enabling multimodal (text, code, data, and image) communication between and within functional groups. Similar to information flow in research laboratories, the ORCHESTRATOR (G1) acts as a central hub, receiving user requests, analyzing requirements, planning, and distributing sub-tasks to administrator agents (function as team leaders) within

each group (G2–G5). These administrator agents further hand off sub-tasks to specialized agents within their respective groups. Leveraging domain expertise and dedicated tools (e.g., GraphRAG for SCIENTIST, Python REPL for ENGINEER, and Matplotlib for ANALYST), these agents execute their assigned tasks. Results are aggregated by the administrators and returned to the ORCHESTRATOR, which evaluates them and adjusts the whole sub-task list, assigning tasks or returning final results. This hierarchical structure, with clearly defined roles and communication pathways, ensures efficient collaboration, robust task management, and a streamlined, adaptable workflow for materials design.

As illustrated in Figure 1C, MARS integrates a human-in-the-loop mechanism, enabling dynamic task adjustment and optimization through user interaction and feedback with the ORCHESTRATOR. This allows expert intervention and validation at critical decision points when agents request assistance from experts, effectively combining computational efficiency with human expertise to ensure practical validity while maintaining autonomous operation. Figure 1D presents the user-friendly interface that facilitates user interaction and real-time monitoring of the experimental workflow (see Note S2 and Figures S1 and S2). As shown in Figure 1E, the capabilities of MARS have been validated by designing and synthesizing

diverse nanomaterials, demonstrating the effectiveness in practical research scenarios.

Knowledge-driven automated experimentation

Retrieval-augmented generation (RAG) enhances language model outputs by incorporating relevant information from external knowledge sources before generating responses. This approach grounds model outputs in factual, domain-specific information, thereby reducing hallucinations that commonly occur when models rely exclusively on knowledge embedded during training.³² MARS integrates materials knowledge retrieval with experimental execution, forming complementary cognitive and physical dimensions that enable the system to transition from conceptual understanding to real-world synthesis. Research on materials typically begins with comprehensive literature review and information retrieval. The SCIENTIST group accelerates this process through a hybrid RAG approach that transforms unstructured scientific literature into structured experimental guidance. Our hybrid RAG framework combines traditional vector similarity search with graph-based retrieval methods (GraphRAG), enabling more comprehensive and contextually aware information retrieval for complex materials science queries.

As illustrated in Figure 2A, four specialized LLM-based agents within the SCIENTIST group focus on distinct areas: structure, synthesis, properties, and applications, each contributing distinct expertise while leveraging specialized retrieval tools. The core retrieval system combines three primary approaches: KeywordSearch for word-level query matching, VectorSearch for semantic-level similarity-based matching, and GraphRAG for topic-level view through the knowledge graph. This hybrid approach ensures comprehensive query results by balancing sentence matching, paragraph matching, and topic view.

As shown in Figure 2B, MARS utilizes an extensive knowledge base carefully constructed from academic papers, books, and web resources to support this retrieval system (see Note S3). These diverse sources are converted to markdown format for natural segmentation and then enhanced with vector representations for similarity searches and structured into knowledge graphs for conceptual reasoning.^{33–35} Online resources (WebSearch, LitSearch, and DatabaseSearch) supplement the offline knowledge base when required. Upon receiving a sub-task from the ORCHESTRATOR, the admin agent of the SCIENTIST group routes it to the most appropriate specialized agent using tailored prompts. These specialized agents utilize hybrid retrieval tools to acquire relevant domain knowledge when needed, thus enhancing the accuracy of their responses.

To demonstrate the capabilities of MARS, the synthesis of perovskite nanocubes is selected as a representative task. CsPbBr₃ perovskites are promising for optoelectronic applications, but their nanocrystal synthesis is complex, with multiple interdependent factors including precursor selection, ligand engineering, reaction temperature control, and precise timing of various steps. For this demonstration, a specialized knowledge base was constructed from 1,348 high-impact perovskite publications, resulting in a vector database (7,818 segments) and a comprehensive knowledge graph (117,749 nodes and 232,490 edges), which provides a rich foundation for material synthesis

reasoning (see Figures S3 and S4). As shown in Figure 2C, MARS demonstrates superior performance by accurately retrieving precise reagent specifications and procedural details (see Note S4), whereas conventional LLMs like GPT-4o, o3-mini, and Llama 3.1-70B produce hallucinations. Regarding hallucinations, we identified several issues: numerical fabrication (such as providing integer values rather than stoichiometric ratios), incorrect synthesis conditions (inappropriate solvent or precursor selection), omission of critical experimental steps, and disregard for user constraints (see Note S5). MARS mitigates these issues through its multi-agent RAG architecture and addresses the knowledge deficit, which is identified as the fundamental cause of hallucinations.³⁶

As shown in Figure 2D, knowledge is transformed into action through the ENGINEER and EXECUTOR groups, which convert conceptual understanding into physical experimentation. Figure 2E illustrates the ENGINEER group, which serves as the crucial bridge, translating synthesis protocols into machine-executable instructions through four specialized agents: the scheme plotter creates visual flowcharts for human verification, the structural engineer converts protocols to standardized JSON schemas, the software engineer generates specialized code when needed,^{37,38} and the code reviewer ensures operational safety and feasibility (see Note S6). Subsequently, the EXECUTOR group implements these instructions through a self-developed experimental robotic platform⁵ (see Figure S5) that integrates workstation controller, mobile-robot controller, and characterization controller agents communicating with unified control middleware (see Figure S6).

Figure 2F demonstrates this workflow through the automated room temperature synthesis of CsPbBr₃ nanocrystals. Upon receiving the synthesis scheme from the SCIENTIST group, the ENGINEER group transforms the protocol into machine-executable instructions, which are subsequently passed to the EXECUTOR group. The workstation and mobile-robot controller agents transmit these instructions through the scheme transmission API tool to the unified control middleware on the computer workstation, thereby initiating the physical synthesis process. The middleware coordinates a sequence of automated operations: (1) mobile robots locate reagents and consumables from storage areas (no. 1–3); (2) the robots transport reagent bottles to the adding device (no. 4–5); (3) the materials are transferred to the liquid workstation (no. 6) for precise pipetting, stirring, and heating; and (4) the completed samples are transported to the fluorescence spectrometer (no. 7) for characterization. Throughout this process, spectroscopic data are transmitted to the middleware and accessed by the DataCollector agent for analysis.

Knowledge-driven automated experimentation

Parameter optimization is a critical component in materials science research. Conventional machine learning methods generally require extensive experimental data to build optimization models, but data are lacking in actual experimental processes. Figure 3A illustrates how MARS implements a paradigm shift by employing knowledge-driven reasoning rather than relying on data-intensive modeling approaches. By leveraging the powerful reasoning capabilities of LLMs combined with comprehensive materials

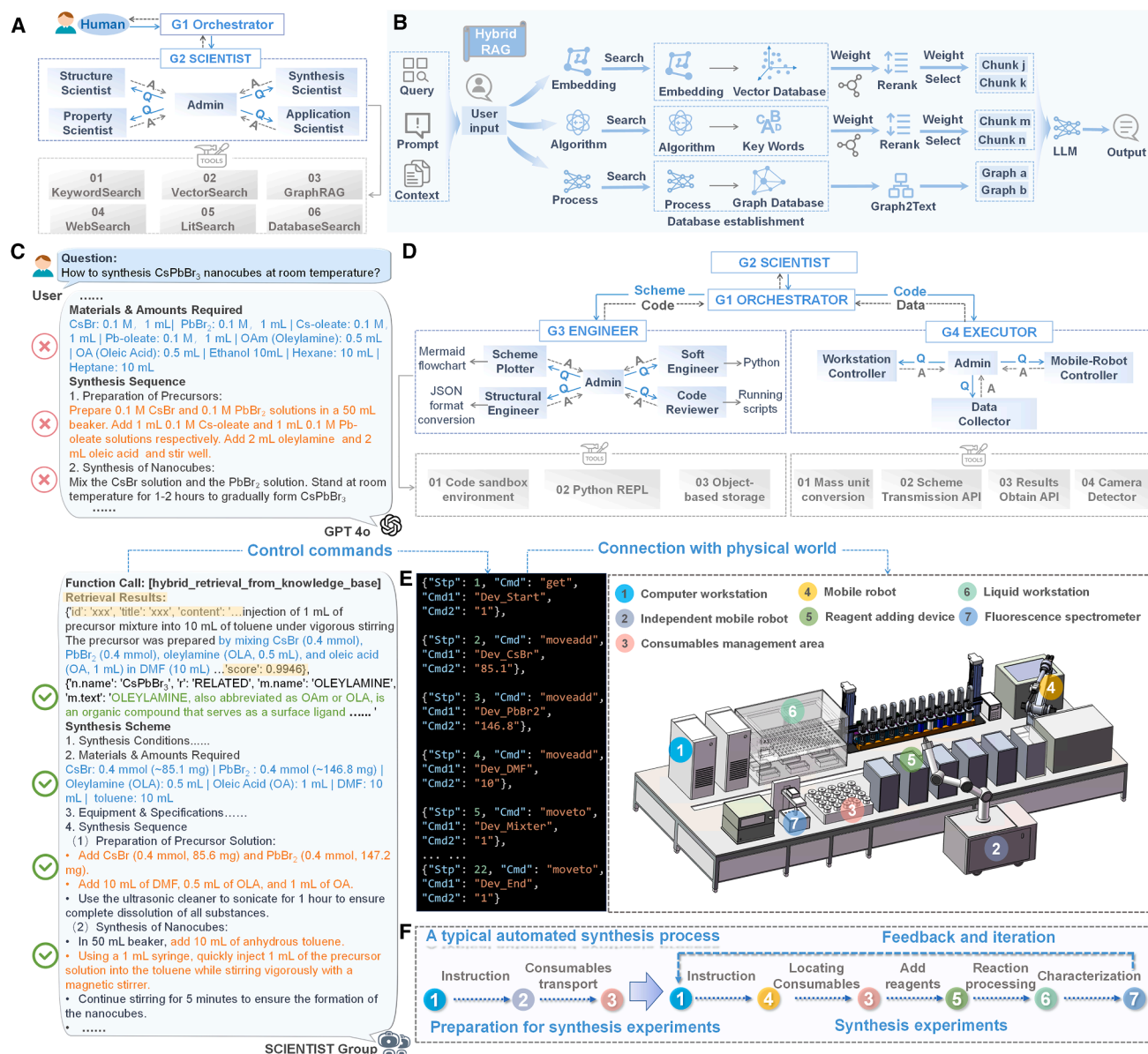


Figure 2. MARS for materials synthesis design and automated experimentation

(A) Overview of the SCIENTIST group.

(B) Schematic diagram of the hybrid RAG approach.

(C) Responses generated by GPT-4o and MARS to the same user query (uniform sampling parameters across systems: top_k = 0, top_p = 1, temperature = 0).

(D) Overview of the ENGINEER and EXECUTOR groups.

(E) Example of code and the controlled experimental robotic platform.

(F) Workflow of typical CsPbBr₃ nanocrystal synthesis process.

knowledge, the system can perform advanced parameter optimization and synthesis prediction with minimal experimental data.¹⁵ This paradigm shift enables efficient exploration of synthesis parameters without the high-throughput experimentation typically required by machine learning methods. The Experiment Optimizer agent demonstrates effective LLM specialization through structured system prompts. The prompt design includes several essential elements: clear role definition ("You are a professional Experiment_Optimizer"), integration with analytical tools (Scikit-

Learn, Optuna, and Matminer for modeling), and access to domain-specific literature through RAG retrieval and online search tools. This design implements a decision framework that directs the agent to select between code-based modeling for data-rich scenarios and knowledge-based parameter adjustment when data are limited, mimicking the actual workflow of materials scientists. Besides, the ANALYST_Admin agent enhances system robustness by providing critical coordination and quality control functions, including verification of chemical validity and

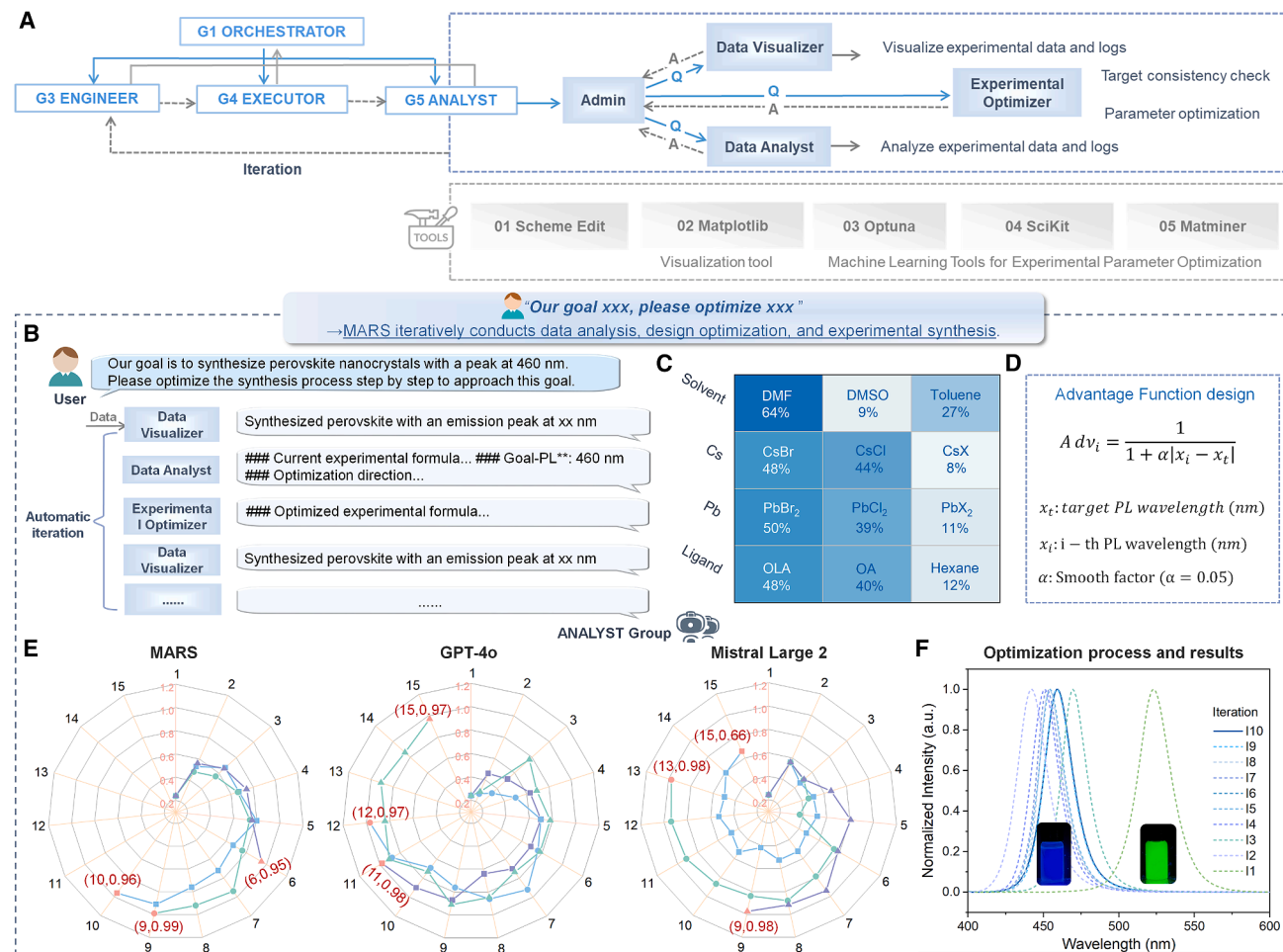


Figure 3. Optimization capabilities of MARS

- (A) Overview of the ANALYST Group.
 (B) Visualization of the multi-round iterative process of the agent.
 (C) Comparison of reagent choices across the knowledge base.
 (D) Mathematical expression used to calculate advantage values.
 (E) Comparison of optimization process performed based on MARS, GPT-4o, and Mistral Large2.
 (F) PL curve in iteration process based on MARS.

identification of duplicate recipes, thereby establishing comprehensive optimization through internal error correction mechanisms that single-model systems typically lack.

As shown in Figure 3B, the capabilities of MARS in feedback-driven optimization were demonstrated through the challenging task of tuning perovskite nanocrystal emission from green (523 nm) to blue (460 nm). All-inorganic perovskite nanocrystals have attracted considerable research interest due to their facile colloidal synthesis, tunable emission wavelengths spanning the entire visible spectrum, high photoluminescence quantum yields, and narrow emission bands. Wavelength tuning through halide composition engineering represents a crucial research direction with significant applications in light-emitting diodes (LEDs), lasers, and quantum light sources. Besides, this task demonstrates a classic materials optimization challenge that re-

quires precise control over quantum confinement effects and composition, typically achieved through halide exchange (Br to Cl substitution) or dimensional control. This optimization process demands professional reasoning about how precursor ratios and reaction conditions affect band-gap engineering in semiconductor nanocrystals, guided by well-defined targets with quantifiable outcomes. From an experimental demonstration perspective, this optimization task provides rapid visual feedback through color changes and enables real-time optical characterization, effectively highlighting the multi-parameter optimization capabilities of MARS.

Figure 3C presents the distribution of key reagents (including solvents, Cs sources, Pb sources, and ligands) reported in the literature for room temperature antisolvent precipitation of blue-emitting perovskites. These statistics were extracted from

the MARS knowledge base to help researchers establish a chemically reasonable design space grounded in prior experimental work. Notably, achieving the target 460 nm emission remains a zero-shot optimization challenge that requires iterative reasoning through complex constraints without prior experimental data for this specific target. Figure 3D shows the merit function, defined as the difference between the emission peak of the product and the target value, which is designed to assess optimization progress. For comparison, three parallel optimization iterations using MARS, currently one of the most powerful closed-source LLMs (GPT-4o), and the leading open-source LLM (Mistral Large2) were performed. To ensure scientific rigor and fairness in our comparative evaluation, we designed a controlled experiment where all participating models (MARS, GPT-4o, and Mistral Large 2) received identical standardized input formats. Notably, in preliminary tests requiring independent synthesis protocol generation for 523 nm CsPbBr₃ nanocrystals (NCs) without any starting recipe, only MARS successfully provided a complete and executable ligand-assisted reprecipitation synthesis protocol. In contrast, GPT-4o and Mistral Large 2 could only provide vague guidance or even demonstrate fundamental chemical misconceptions, highlighting the professional knowledge gap between specialized and general-purpose models.

All models began optimization from the same initial recipe (523 nm emission) provided by MARS for the formal optimization comparison. Optimization concluded upon reaching a merit value >0.95 or after 15 iterations. To ensure fair comparison, the optimization suggestions from GPT-4o and Mistral Large 2 were manually input into the automated experimental platform of MARS for execution, as MARS functions as an integrated design-experiment-optimization system capable of directly executing synthesis protocols and obtaining experimental feedback, whereas other models only provide textual responses. Each model received identical iterative feedback in a standardized table format containing current emission peak, historical data, and optimization targets (see Note S7). As presented in Figure 3E, MARS requires an average of only 8.3 iterations to reach the stopping condition, while GPT-4o and Mistral Large2 require 12.7 and 12.3 iterations, respectively. Mistral Large2 even fails to reason correctly once under the given rules, remaining far from the target even after more than 15 iterations. Figure 3F demonstrates how MARS achieves the target within ten iterative optimization cycles, highlighting its capability to autonomously use domain knowledge for guiding subsequent experimental decisions. This knowledge-driven approach significantly reduces the experimental burden by requiring fewer iterations to reach targets, allowing researchers to explore synthesis parameters more efficiently than standalone LLMs. Superior optimization efficiency in the MARS system is achieved through four key design elements that address fundamental limitations of general-purpose LLMs in materials science applications. First, specialized prompt engineering has proven effective for enhancing domain-specific LLM performance.^{39,40} Second, the multi-agent architecture enables task decomposition and specialization, avoiding cognitive overload when single models handle heterogeneous tasks.^{41,42} Third, structured decision flows with internal quality checks provide multiple error

correction points, ensuring reliable outputs.⁴³ Finally, the knowledge-enhanced mechanism through hybrid RAG provides professional depth far exceeding general models, enabling scientifically grounded optimization strategies rather than blind trial and error.^{44,45}

Knowledge-driven materials design

The design of novel functional materials typically requires researchers to possess extensive cross-disciplinary knowledge and rich experiential accumulation, particularly for complex material systems that necessitate balancing multiple performance dimensions. Figure 4 demonstrates in detail how the MARS system overcomes limitations of traditional development paradigms to assist scientists in achieving materials design and synthesis validation of perovskite ternary nanocomposite structures (see Note S8). As shown in Figure 4A, traditional materials design cycles typically require several months, encompassing literature review, experimental design, materials synthesis, result validation, and experimental optimization. This process heavily depends on researchers' efforts and experience while simultaneously constraining innovation possibilities due to insufficient systematic knowledge integration. In contrast, the MARS-assisted design cycle accelerates this process significantly and enables efficient material design through knowledge-driven and collaborative operations of multiple agents (G1–G5), compressing the entire procedure to just a few hours.

While precise wavelength control demonstrates the optimization capabilities, the fundamental challenge limiting perovskite nanocrystal applications is their poor stability. Lead halide perovskites like CsPbBr₃ possess ionic crystal structures that readily decompose upon water exposure, causing complete fluorescence quenching within hours. This weakness prevents practical deployment in photovoltaics, LEDs, and other optoelectrical applications. According to Web of Science analysis (see Figure S7), research on perovskite water stability has grown exponentially, reflecting the urgent need for effective stabilization strategies. Given this critical challenge, we tasked MARS with designing water-stable perovskite nanocrystals to showcase its capability in addressing pressing materials science problems.

The multi-round interaction process illustrated in Figure 4B reveals how the MARS progresses from mechanistic understanding to creative design through increasingly refined problem analysis and solution strategies. In round 1, when researchers inquired about core-shell stabilization mechanisms, MARS provided comprehensive explanations of how this approach protects perovskites through moisture barriers, defect passivation, thermal stability enhancement, and ion migration suppression. During this exploration of stability solutions, researchers noted an additional opportunity revealed by literature analysis: the growing interest in biomedical applications of perovskite materials (see Figure S8). These applications require both stability and biocompatibility as essential properties. This insight guided the subsequent interaction in round 2, where researchers introduced biocompatibility as an additional requirement. MARS identified this dual challenge and proposed multiple candidate solutions, including oxides, polymers, and biomimetic shells, each with distinct advantages for different application scenarios.

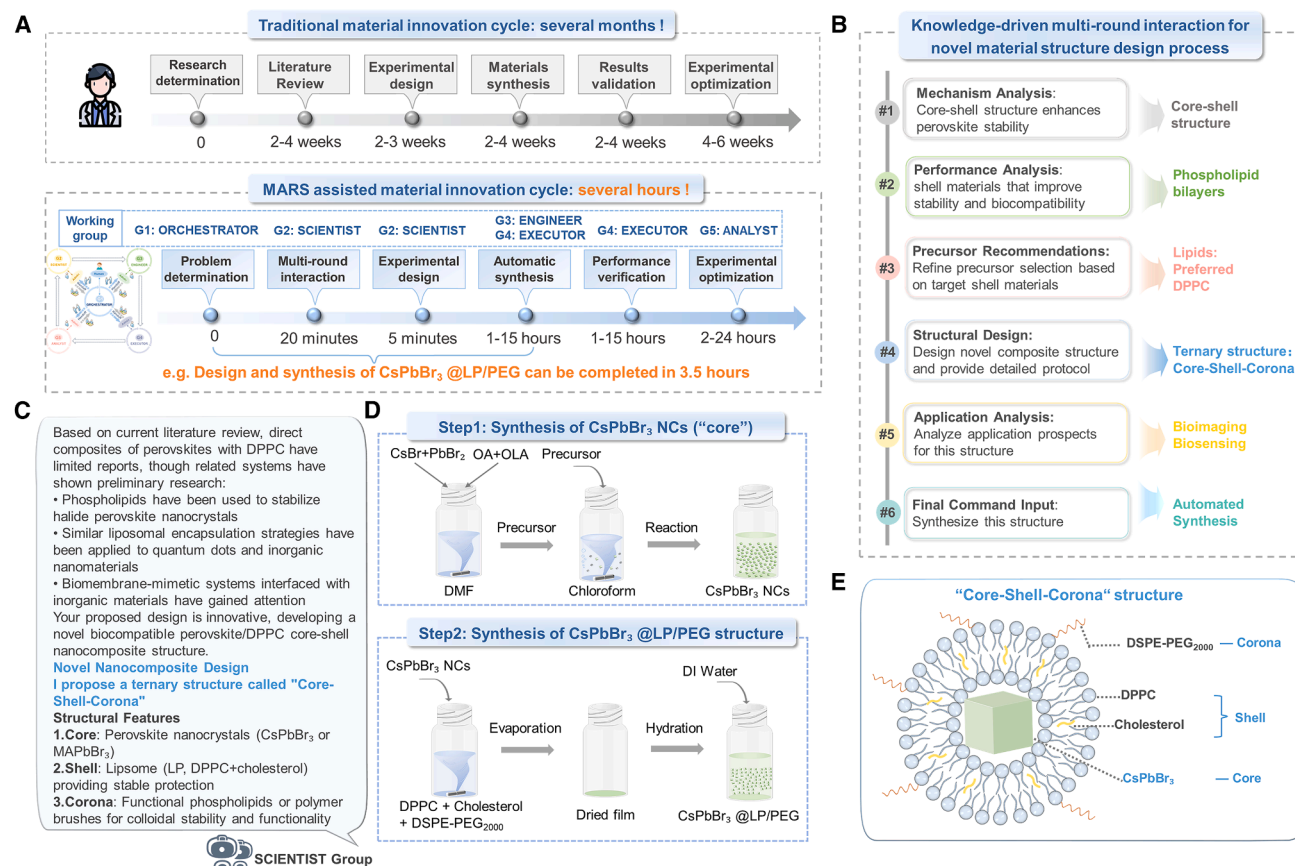


Figure 4. MARS-assisted design and synthesis of perovskite nanocomposite materials

(A) Efficiency comparison between the traditional materials design cycle and the MARS-assisted design process.

(B) Multi-round knowledge-driven interaction process for designing CsPbBr_3 @LP/PEG composite materials.

(C) Biomimetic membrane ternary "core-shell-corona" design.

(D) Synthesis route for CsPbBr_3 @LP/PEG.

(E) Structural illustration of CsPbBr_3 @LP/PEG.

From these options, phospholipid encapsulation was ultimately selected by researchers due to two key advantages: the amphiphilic nature of phospholipids facilitates self-assembly on nanocrystal surfaces under mild conditions that preserve crystal structure while providing protective encapsulation; and as major components of biological membranes, phospholipids inherently offer excellent biocompatibility. This strategic selection resulted in materials possessing the combined properties of excellent fluorescence, environmental stability, and biocompatibility required for advanced applications in biosensing, cell imaging, and tumor-targeted therapy.

When researchers specifically asked about phospholipid bilayer encapsulation in round 3, MARS provided detailed precursor recommendations including various phospholipids, such as 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC), 1,2-Dioleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (DOPG), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), and functional components (cholesterol and 1,2-Distearoyl-sn-glycero-3-phosphorylethanolamine [DSPE]). Following this prioritized recom-

mendation, researchers specified DPPC for further design in round 4. As shown in Figure 4C, MARS identified limited research reports on direct perovskite-DPPC composites through its retrieval results and then analyzed relevant advances in liposomes, quantum dot encapsulation, and biomimetic membrane systems to establish foundations for innovative design. The system proposed a "core-shell-corona" ternary structure concept, utilizing a lipid bilayer to provide hydrophobic barriers while introducing a polyethylene glycol (PEG) corona to enhance colloidal stability. Specifically, the resulting ternary structure includes (1) core: CsPbBr_3 or MAPbBr_3 perovskite nanocrystals, providing excellent optical properties; (2) shell: a liposomal structure composed of DPPC and cholesterol, with cholesterol incorporation helping to modulate membrane fluidity and mechanical stability, forming a dense barrier layer that blocks water and oxygen molecule penetration; and (3) corona: DSPE-PEG₂₀₀₀ functional polymer brushes, providing colloidal stability through steric repulsion while conferring biocompatibility and potential for targeted functionalization. Notably, MARS prioritized DPPC in its phospholipid recommendations over the commonly used

DOPC in the literature.^{46–49} The scientific rationale for this recommendation becomes clear when comparing their molecular structures: although both DPPC and DOPC possess neutral phosphocholine headgroups, the saturated palmitic chains (16:0) of DPPC form more densely packed membranes with significantly lower water permeability compared to the unsaturated oleic chains (18:1) of DOPC.⁵⁰ The incorporation of cholesterol further enhances membrane organization by promoting liquid-ordered phase formation with DPPC, resulting in an exceptionally robust water barrier.^{51,52} This multi-level structural design achieves synergistic optimization of perovskite surface defect passivation, physical barriers against water/oxygen permeation, and colloidal stability at the molecular level, demonstrating the capability of MARS to integrate multidisciplinary knowledge for creative design.

Figure 4D illustrates the two-step synthesis route designed by MARS, reflecting its professional translation capability from theoretical design to experimental protocol (see Note S8). After the researchers acknowledged this design proposal and issued synthesis instructions, MARS generated JSON protocols and automatically performed the synthesis of this composite on our robotic platform. Within 3.5 h of the initial query, perovskite nanocrystal aqueous solutions were successfully obtained (Figure 4E). To investigate the behavior of CsPbBr₃@LP/PEG NCs in an aqueous environment, the evolution of their photoluminescence (PL) was tracked. As detailed in Figure S9A, the PL emission exhibited an initial slow increase followed by a gradual decrease upon water immersion. This fluorescence evolution is not uncommon and has been observed in other phospholipid-coated CsPbX₃ systems.⁵³ Previous studies have demonstrated that interactions with water can lead to surface defect passivation and induce a phase transition from CsPbBr₃ to CsPb₂Br₅, forming heterostructures that initially enhance fluorescence.^{54,55} Our X-ray crystallography (XRD) characterization confirmed the formation of the CsPb₂Br₅ phase during water immersion, consistent with these literature reports (see Figure S9B). Cytotoxicity testing confirmed the effectiveness of phospholipid encapsulation in enhancing perovskite biocompatibility. In the evaluation, CsPbBr₃@LP/PEG maintained >95% cell viability at concentrations up to 100 µg/mL, while pristine CsPbBr₃ showed significant cytotoxicity (see Figure S10). These results demonstrate that phospholipid encapsulation effectively prevents lead ion leaching while improving optical stability for further biological applications.

Performance evaluation of the MARS framework

The capabilities of MARS were rigorously assessed through pairwise win rate based on established LLM evaluation frameworks (see Note S11).⁵⁶ Specifically, a benchmark dataset comprising 929 materials science text fragments with standardized test questions was constructed through an automated question-answer pair generation protocol. Claude3.5-Sonnet was employed to conduct iterative pairwise comparisons between model responses, with randomized response ordering to mitigate positional bias.⁵⁷ This method significantly enhances the measurability of subtle differences between models, ensuring robust discriminative validity and statistical reliability.

As shown in Figure 5A, the fundamental architectural advantage of MARS is demonstrated through systematic comparisons

using identical base models. When GPT-4o is employed as the foundation in both systems, MARS achieves a 60.02% win rate against the single-agent implementation. This significant performance reveals critical architectural differences: MARS overcomes the limitations of single-agent systems through hierarchical design, where performance typically degrades when processing extensive contexts and managing multiple simultaneous tasks. By implementing specialized agents with defined responsibilities and enabling multi-round, depth-focused information retrieval within isolated task groups, MARS effectively reduces cognitive burden on individual models while enhancing overall system stability and knowledge integration capabilities.³⁶ Additionally, MARS also outperforms ChatGPT-4o (the web interface version), achieving a 61.89% win rate. This result demonstrates how domain-specific knowledge integration and hierarchical multi-agent architecture provide advantages over even highly optimized consumer AI systems.

Figure 5B demonstrates the adaptability of MARS across different foundation models, validating the broad applicability of the framework. Beyond the closed-source GPT-4o accessed via API, MARS was successfully implemented using DeepSeek-R1, a powerful open-source model with strong reasoning capabilities. The pairwise comparison results show that MARS (R1) consistently outperforms both the single-agent (R1) and the standalone DeepSeek-R1 model, confirming that the multi-agent architecture provides performance advantages regardless of foundation models. The average win rate analysis in Figure 5C reveals that MARS systems achieve higher overall performance compared to their single-agent counterparts and base models. The overall win rates were calculated by averaging win rates from all unique, non-self pairwise comparisons (see Figure S11). The performance hierarchy clearly shows that MARS benefits from stronger foundation models, with the R1-based implementation achieving higher average win rates than the GPT-4o version. This dependence on base model capabilities suggests that as LLMs continue to improve, the MARS framework will correspondingly enhance its performance in materials science applications. The multi-agent architecture thus provides a scalable approach that can adapt to and benefit from ongoing advances in LLMs.

While automated benchmarking provides quantitative comparisons, validating the practical utility of MARS in domain-specific scenarios requires expert assessment. Our evaluation panel comprised 10 researchers with extensive expertise in nanomaterial synthesis: 3 PhD candidates, 3 assistant researchers, 2 associate researchers, and 2 professors. All evaluations followed a double-blind protocol, with each test case independently scored by 5 experts. We employed a 10-point comprehensive scoring system evaluating three primary dimensions: scientific accuracy (40%), task completion (30%), and reasoning quality (30%). The evaluation tasks spanned four key aspects of materials research: structure design, synthesis protocol development, property prediction, and application optimization. Detailed results and task-specific performance analyses are provided in Note S9, where MARS consistently demonstrated superior accuracy and experimental actionability compared to baseline models (Figure S12).

Having demonstrated the effectiveness of MARS in perovskite nanocrystal synthesis, optimization, and design in the

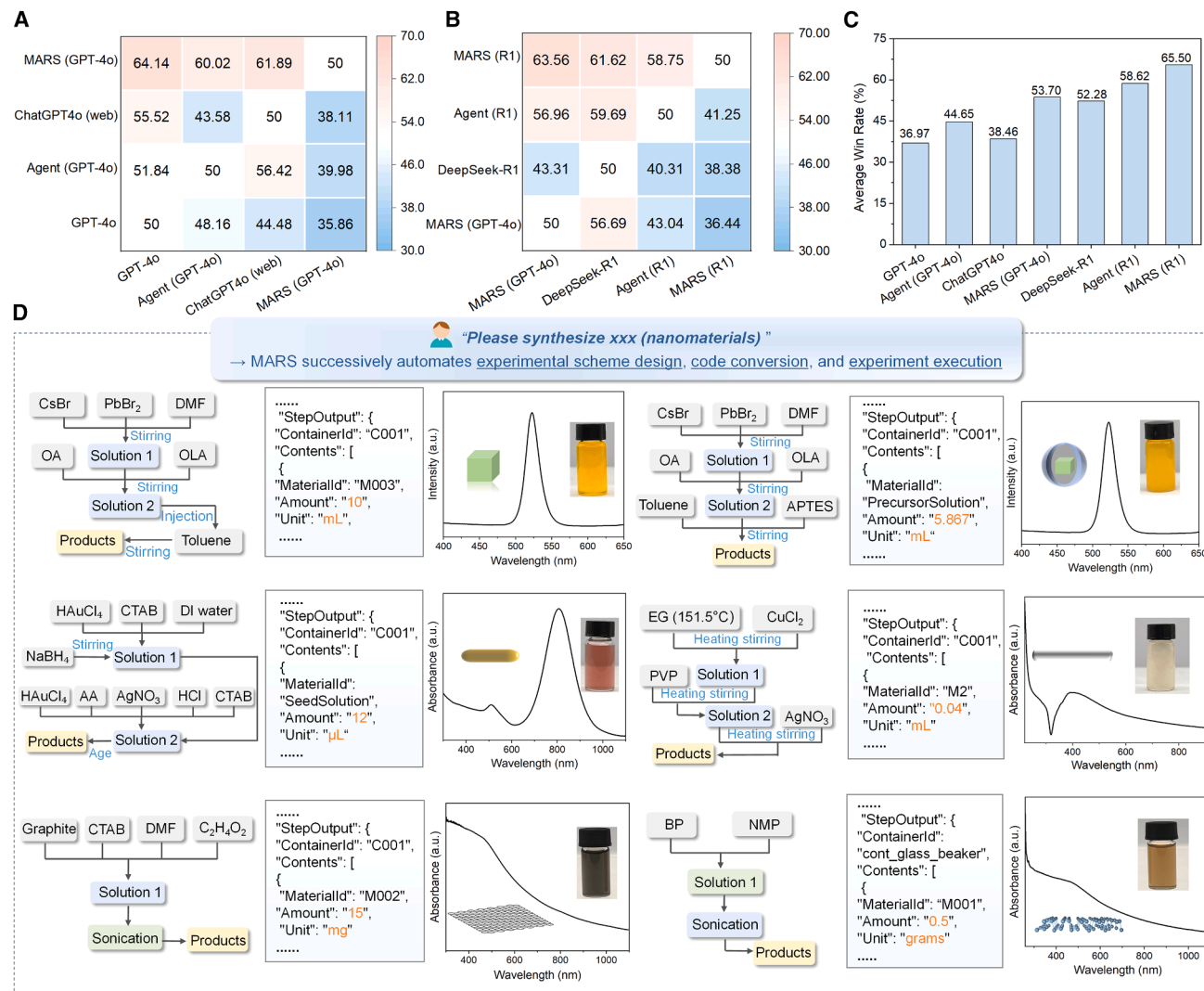


Figure 5. Evaluation and practical demonstration of MARS

(A) Pairwise win rate between MARS and single-agent (GPT-4o), ChatGPT-4o web interface, and GPT-4o (50% win rate for self-comparison, not tested).

(B) Pairwise win rate matrix demonstrating framework scalability (50% win rate for self-comparison, not tested).

(C) Average win rates across all pairwise comparisons.

(D) Validation of the automated synthesis capabilities of MARS across diverse nanomaterial systems.

main study, the evaluation was further extended to encompass a broader range of liquid-phase nanomaterial systems. As demonstrated in Figure 5D, MARS was successfully applied to materials across different dimensionalities, including Au nanorods, Ag nanowires, graphene, and black phosphorus nanosheets. For each material system, MARS autonomously designed natural language synthesis protocols, generated structured JSON protocols, and executed experimental procedures. This comprehensive validation across multiple evaluation methods and diverse material systems, combined with the proven adaptability of the framework to different foundation models, establishes MARS as a broadly applicable platform for accelerating materials research through intelligent automation.

DISCUSSION

This work introduces MARS, an LLM-driven platform for autonomous materials development. MARS integrates 19 specialized agents within a hierarchical framework, which are organized into functional groups to facilitate collaborative execution of complex tasks spanning design, synthesis, and optimization. By leveraging a curated knowledge base and a hybrid RAG approach, MARS achieves enhanced accuracy and mitigates LLM hallucinations, enabling professional-grade, autonomous experimentation. Significantly, MARS achieved rapid optimization of material properties (e.g., tuning green-emitting perovskite nanocrystals to blue emission within 10 iterations) and discovery of novel materials within hours, significantly accelerating the

research process. Beyond efficiency gains, the MARS architecture enables seamless information flow and knowledge accumulation throughout the research pipeline, creating a responsive closed-loop system that adapts to experimental feedback in real time.

MARS represents a significant step toward autonomous materials research. This system reduces substantial repetitive work, allowing researchers to focus on creative thinking and scientific insights. The flexible, collaborative agent framework permits customization by researchers, enabling integration of local databases and adaptation to diverse experimental platforms. Such features facilitate tailoring to specific research needs while expanding applicability across various materials and synthesis techniques. Despite these advantages, current limitations must be acknowledged. MARS primarily relies on existing literature knowledge for combinatorial optimization, with its innovative capacity currently limited to incremental improvements rather than breakthrough discoveries. This reflects fundamental challenges in applying LLMs to scientific research: distinguishing rational inference from hallucinatory outputs and balancing knowledge dependence with exploratory innovation. The system functions as an intelligent research assistant that accelerates discovery cycles rather than replacing human creativity.

Beyond performance, reproducibility and runtime error handling represent critical factors for practical adoption in laboratory settings. The MARS architecture addresses reproducibility through a comprehensive open-source implementation with standardized interfaces between agent modules and robotic components. This design allows other laboratories to replicate system functionality, though some adaptation to specific hardware configurations remains necessary. Regarding error management, MARS employs a multi-tiered approach to handle failures in both design and synthesis processes: human intervention is strategically integrated at critical decision points, autonomous replanning occurs when agents detect suboptimal outcomes, and rule-based verification protects robotic execution. This flexible error handling capability ensures experimental continuity despite the unpredictable challenges inherent in materials research, maintaining system reliability without constant expert supervision.

Future development will focus on standardized interfaces for laboratory equipment, enabling more complex experimental control. Beyond predefined functions, future LLMs could autonomously devise experimental strategies, optimize parameters, and adapt to unexpected situations, realizing embodied intelligence in materials research. Besides, while the hybrid RAG approach mitigates hallucinations, further advancements in LLM architectures and domain-specific fine-tuning are crucial. Critical directions include integrating physics-based simulations and first-principles calculations to enhance scientific reasoning capabilities, developing more reliable metrics for assessing innovation versus hallucination. We believe the evolution from knowledge-driven to genuinely autonomous innovation represents the necessary trajectory for AI in scientific research, with the current work establishing an important foundation for this long-term goal. Future MARS iterations promise to accelerate materials discovery and development to meet evolving scientific and industrial demands.

METHODS

Architecture details

MARS design

The MARS system utilizes GPT-4o from OpenAI as its foundation model, with the temperature parameter set to 0.1 to ensure high stability and consistency during task execution. The AutoGen framework (<https://github.com/microsoft/autogen>) simplifies complex LLM workflow orchestration, automation, and optimization through multi-agent dialogue mechanisms. Its flexible architecture enables developers to customize agent behaviors and interaction patterns according to specific task requirements. AutoGen provides two core agent types: AssistantAgent and UserProxyAgent. In this study, we utilize AssistantAgent provided by AutoGen, defining agent roles and functions through system_message and tools parameters. Simultaneously, we set up UserProxyAgent as a human control interface to implement human-in-the-loop in the multi-agent system, enabling human intervention at critical points during task execution. In terms of agent orchestration, MARS primarily employs two modes: Selector Group Chat and Swarm (<https://github.com/openai/swarm>). Selector Group Chat is a multi-agent dialogue mode based on model selection mechanisms, where agents broadcast messages to other members in turns, and an Orchestrator agent selects the next speaking agent or team based on conversation context. The Swarm mode is a multi-agent collaboration mode based on task delegation mechanisms. Agents dynamically delegate tasks to more suitable agents based on task requirements and capabilities, ensuring efficient task allocation and execution within the team. The core mechanism of Swarm relies on agents generating HandoffMessages to specify task delegation targets. The admin agents between teams coordinate through the Selector Group Chat mode, while task allocation within teams relies on the Swarm mode. Further details regarding the methods can be found in the supplemental information.

SCIENTIST group

The scientist team consists of four materials science agents, each initialized through AssistantAgent. We guide different scientist agents to form different capability preferences through system prompts. Within the scientist team, we use AutoGen's Swarm mode to orchestrate agents, with four scientist agents communicating directly with the Scientist Team Admin agent through task submission.

Knowledge base and retrieval tools

We established a knowledge base containing over 100,000 nano-material-related literature papers and developed interfaces based on keyword retrieval and vector retrieval for efficient research material access. Specifically, we used the MinerU tool to convert literature PDFs to markdown format for an efficient knowledge base. For the SCIENTIST group, we customized the following retrieval tools. *Vector database and search*: to optimize the RAG system's retrieval, we segmented papers by section and vectorized text segments using the gte-Qwen2-1.5B-instruct model (<https://huggingface.co/thenlper/gte-large>). The vectors are stored in a Weaviate database of Dify, with API functionalized as hybrid_retrieval_from_knowledge_base providing KeywordSearch and VectorSearch interfaces based on

keywords and vectors. *Knowledge graph*: based on the knowledge base, we classified literature by research topics and utilized GraphRAG to construct knowledge graphs for different research topics, with corresponding interfaces available through Autogen Extension. *WebSearch and LitSearch*: we integrated WebSearch tools based on the Bing search engine and LitSearch tools based on Arxiv and Semantic Scholar for real-time online searches. Bing searches, arXiv, and Semantic Scholar are called directly through official APIs, which functioned as “paper_retrieval_from_arxiv_and_pubmed” and “online_retrieval_with_bing_and_playwright.” *DatabaseSearch*: material databases are crucial in nano-materials research. We developed database search tools based on OQMD and Materials Project (MP) databases, providing comprehensive and precise material data query capabilities. While the MP database offers official retrieval interfaces, we simplified and encapsulated their parameters through FastAPI to prevent agent call failures. For the OQMD database, we obtain query results using PlayWright and similarly encapsulate them using FastAPI.

ENGINEER group

The MARS engineer agent team consists of four agents defined based on AutoGen’s AssistantAgent class, coordinating through system prompts and tools. Within the team, we employ AutoGen’s Swarm mode to organize engineer agents and admin agents, efficiently completing subtasks distributed by the Orchestrator through task submission. The following tools support the ENGINEER group. *Code Sandbox Environment*: AutoGen provides DockerCommandLineCodeExecutor, which relies on Docker container isolation environments for code execution that might cause conflicts or crashes in local Python environments. *Python REPL*: we directly use PythonCodeExecutionTool to execute Python code reviewed and confirmed safe by the code reviewer. *Object-based storage*: we introduced object storage tools using MinIO to build a local object storage service, encapsulated through FastAPI, to maintain the integrity and stability of structured experimental schemes.

EXECUTOR group

To address potential process blocking during long wait times for characterization data collection, data collection tools are set with a 6-h execution time and a 30-s sleep mechanism when data are unavailable. We customized the following tools for the EXECUTOR group. *Mass unit conversion* implemented the mol2mg tool based on the periodictable Python package for converting molar mass to weight units. *Scheme transmission API*: implemented using FastAPI for transmitting structured synthesis scheme access links directly to robots or island stations. *Results obtain API*: implemented using FastAPI for direct access to characterization device data. *Camera detector* provides IP Camera video stream interface based on RTSP protocol for real-time monitoring of mobile robots and island stations.

Experimental details

Cell viability assessment

MCF-7 cells were seeded in 96-well plates (1×10^4 cells/well) and cultured for 24 h. Cells were then incubated with various concentrations (0–100 $\mu\text{g/mL}$) of pristine CsPbBr₃ or CsPbBr₃/LP/PEG for 24 h. CCK-8 reagent was added, and absorbance was

measured at 450 nm. Live/dead staining was performed using the Calcein-AM/PI kit following the manufacturer’s protocol.

Viable and dead cell staining

MCF-7 cells were incubated in 6-well plates under standard conditions for 24 h. After removal of the digested medium, the cells were treated with fresh medium that contained pristine CsPbBr₃ or CsPbBr₃/LP/PEG for 4 h of co-culture. Then, cells were stained with calcein AM (10 μM) and PI (10 μM) according to the product description for 0.5 h. The coloring solution was removed and rinsed twice gently with cold PBS. Finally, the cells were imaged using a fluorescence inverted microscope (Olympus UHGLGPS, China).

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Xue-Feng Yu (xf.yu@siat.ac.cn).

Materials availability

Samples can be generated following the MARS workflow as described in the methods.

Data and code availability

The data supporting the findings of this study are available within the paper and in the supplemental information files. Source data and Python scripts for MARS developed in this study are available at GitHub: <https://github.com/tangger2000/MARS>.

ACKNOWLEDGMENTS

This work was financially supported by the National Key R&D Program of China (2023YFA0915600), Shenzhen Medical Research Fund (B2302028), National Natural Science Foundation of China (32471459), Strategic Priority Research Program of the Chinese Academy of Sciences (XDB0930000), Shenzhen Science and Technology Program Grant (RCJC20200714114435061, ZDSYS20220527171406014, and KJZD20230923114703007), and the post-doctoral fellowship program of CPSF under grant number GZC20241848.

AUTHOR CONTRIBUTIONS

Conceptualization, T.S. and X.-F.Y.; methodology, T.S., Y.L., Z.W., W.X., G.J., and J.Z.; investigation, T.S., Y.L., and H.H.; visualization, T.S.; supervision, X.-F.Y.; writing – original draft, T.S., Y.L., and Z.W.; writing – review and editing, G.J., D.D., Z.W., R.H., S.R., P.K.C., W.Z., and X.-F.Y.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.matt.2025.102577>.

Received: May 22, 2025

Revised: August 22, 2025

Accepted: November 24, 2025

Published: January 21, 2026

REFERENCES

1. Rothmund, P., Kim, Y., Heisser, R.H., Zhao, X., Shepherd, R.F., and Keplinger, C. (2021). Shaping the future of robotics through materials

- innovation. *Nat. Mater.* 20, 1582–1587. <https://doi.org/10.1038/s41563-021-01158-1>.
2. Mohanty, A.K., Vivekanandhan, S., Pin, J.-M., and Misra, M. (2018). Composites from renewable and sustainable resources: Challenges and innovations. *Science* 362, 536–542. <https://doi.org/10.1126/science.aat9072>.
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. (2020). A mobile robotic chemist. *Nature* 583, 237–241. <https://doi.org/10.1038/s41586-020-2442-2>.
4. Angelone, D., Hammer, A.J.S., Rohrbach, S., Krambeck, S., Granda, J.M., Wolf, J., Zaleskiy, S., Chisholm, G., and Cronin, L. (2021). Convergence of multiple synthetic paradigms in a universally programmable chemical synthesis machine. *Nat. Chem.* 13, 63–69. <https://doi.org/10.1038/s41557-020-00596-9>.
5. Zhao, H., Chen, W., Huang, H., Sun, Z., Chen, Z., Wu, L., Zhang, B., Lai, F., Wang, Z., Adam, M.L., et al. (2023). A robotic platform for the synthesis of colloidal nanocrystals. *Nat. Synth.* 2, 505–514. <https://doi.org/10.1038/s44160-023-00250-5>.
6. Liz-Marzán, L.M., Kagan, C.R., and Millstone, J.E. (2020). Reproducibility in nanocrystal synthesis? Watch out for impurities. *ACS Nano* 14, 6359–6361. <https://doi.org/10.1021/acsnano.0c04709>.
7. Yao, Y., Dong, Q., Brozena, A., Luo, J., Miao, J., Chi, M., Wang, C., Kevrekidis, I.G., Ren, Z.J., Greeley, J., et al. (2022). High-entropy nanoparticles: Synthesis-structure-property relationships and data-driven discovery. *Science* 376, eabn3103. <https://doi.org/10.1126/science.abn3103>.
8. Lu, X.-G. (2015). Remarks on the recent progress of Materials Genome Initiative. *Sci. Bull.* 60, 1966–1968. <https://doi.org/10.1007/s11434-015-0937-2>.
9. Eng, A.Y.S., Soni, C.B., Lum, Y., Khoo, E., Yao, Z., Vineeth, S.K., Kumar, V., Lu, J., Johnson, C.S., Wolverton, C., and Seh, Z.W. (2022). Theory-guided experimental design in battery materials research. *Sci. Adv.* 8, eabm2422. <https://doi.org/10.1126/sciadv.abm2422>.
10. Urbina, F., Lentzos, F., Invernizzi, C., and Ekins, S. (2022). Dual use of artificial-intelligence-powered drug discovery. *Nat. Mach. Intell.* 4, 189–191. <https://doi.org/10.1038/s42256-022-00465-9>.
11. van Dis, E.A.M., Bollen, J., Zuidema, W., Van Rooij, R., and Bockting, C.L. (2023). ChatGPT: five priorities for research. *Nature* 614, 224–226. <https://doi.org/10.1038/d41586-023-00288-7>.
12. Zheng, Z., Zhang, O., Nguyen, H.L., Rampal, N., Alawadhi, A.H., Rong, Z., Head-Gordon, T., Borgs, C., Chayes, J.T., and Yaghi, O.M. (2023). ChatGPT research group for optimizing the crystallinity of MOFs and COFs. *ACS Cent. Sci.* 9, 2161–2170. <https://doi.org/10.1021/acscentsci.3c01087>.
13. Zheng, Z., Zhang, O., Borgs, C., Chayes, J.T., and Yaghi, O.M. (2023). ChatGPT chemistry assistant for text mining and the prediction of MOF Synthesis. *J. Am. Chem. Soc.* 145, 18048–18062. <https://doi.org/10.1021/jacs.3c05819>.
14. Zheng, Z., Rong, Z., Rampal, N., Borgs, C., Chayes, J.T., and Yaghi, O.M. (2023). A GPT-4 reticular chemist for guiding MOF discovery. *Angew. Chem. Int. Ed.* 62, e202311983. <https://doi.org/10.1002/anie.202311983>.
15. Kim, S., Jung, Y., and Schrier, J. (2024). Large language models for inorganic synthesis predictions. *J. Am. Chem. Soc.* 146, 19654–19659. <https://doi.org/10.1021/jacs.4c05840>.
16. Ridnik, T., Kredo, D., and Friedman, I. (2024). Code generation with alpha-codium: From prompt engineering to flow engineering. preprint at arXiv. <https://doi.org/10.48550/arXiv.2401.08500>.
17. Boiko, D.A., MacKnight, R., Kline, B., and Gomes, G. (2023). Autonomous chemical research with large language models. *Nature* 624, 570–578. <https://doi.org/10.1038/s41586-023-06792-0>.
18. Microsoft Research AI4Science, Microsoft Azure Quantum (2023). The impact of large language models on scientific discovery: a preliminary study using GPT-4. preprint at arXiv. <https://doi.org/10.48550/arXiv.2311.07361>.
19. Wahab, M.F., Handlovic, T.T., Roy, S., Burk, R.J., and Armstrong, D.W. (2024). Solving advanced task-specific problems in measurement sciences with generative AI. *Anal. Chem.* 96, 12387–12394. <https://doi.org/10.1021/acs.analchem.4c01734>.
20. M Bran, A., Cox, S., Schilter, O., Baldassari, C., White, A.D., and Schwalter, P. (2024). Augmenting large language models with chemistry tools. *Nat. Mach. Intell.* 6, 525–535. <https://doi.org/10.1038/s42256-024-00832-8>.
21. Bensaude-Vincent, B., and Hessebruch, A. (2004). Materials science: a field about to explode? *Nat. Mater.* 3, 345–347. <https://doi.org/10.1038/nmat1140>.
22. 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. (2023). An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* 624, 86–91. <https://doi.org/10.1038/s41586-023-06734-w>.
23. Yoshikawa, N., Skreta, M., Darvish, K., Arellano-Rubach, S., Ji, Z., Bjørn Kristensen, L., Li, A.Z., Zhao, Y., Xu, H., Kuramshin, A., et al. (2023). Large language models for chemistry robotics. *Auton. Robots* 47, 1057–1086. <https://doi.org/10.1007/s10514-023-10136-2>.
24. Song, T., Luo, M., Zhang, X., Chen, L., Huang, Y., Cao, J., Zhu, Q., Liu, D., Zhang, B., Zou, G., et al. (2025). A multi-agent-driven robotic AI chemist enabling autonomous chemical research on demand. *J. Am. Chem. Soc.* 147, 12534–12545. <https://doi.org/10.1021/jacs.4c17738>.
25. Tom, G., Schmid, S.P., Baird, S.G., Cao, Y., Darvish, K., Hao, H., Lo, S., Pablo-García, S., Rajaonson, E.M., Skreta, M., et al. (2024). Self-driving laboratories for chemistry and materials science. *Chem. Rev.* 124, 9633–9732. <https://doi.org/10.1021/acs.chemrev.4c00055>.
26. Ruan, Y., Lu, C., Xu, N., Zhang, J., Xuan, J., Pan, J., Fang, Q., Gao, H., Shen, X., Ye, N., et al. (2024). Accelerated end-to-end chemical synthesis development with large language models. preprint at ChemRxiv. <https://doi.org/10.26434/chemrxiv-2024-6wmq4>.
27. Chen, K., Lu, J., Li, J., Yang, X., Du, Y., Wang, K., Shi, Q., Yu, J., Li, L., Qiu, J., et al. (2023). Chemist-X: Large language model-empowered agent for reaction condition recommendation in chemical synthesis. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2311.1077>.
28. Sanchez, C., Arribart, H., and Guille, M.M.G. (2005). Biomimetic and bioinspiration as tools for the design of innovative materials and systems. *Nat. Mater.* 4, 277–288. <https://doi.org/10.1038/nmat1349>.
29. Marchese Robinson, R.L., Lynch, I., Peijnenburg, W., Rumble, J., Klaessig, F., Marquardt, C., Rauscher, H., Puzyn, T., Purian, R., Åberg, C., et al. (2016). How should the completeness and quality of curated nanomaterial data be evaluated? *Nanoscale* 8, 9919–9943. <https://doi.org/10.1039/C5NR08944A>.
30. Comandella, D., Gottardo, S., Río-Echevarría, I.M., and Rauscher, H. (2020). Quality of physicochemical data on nanomaterials: an assessment of data completeness and variability. *Nanoscale* 12, 4695–4708. <https://doi.org/10.1039/C9NR08323E>.
31. Darvish, K., Skreta, M., Zhao, Y., Yoshikawa, N., Som, S., Bogdanovic, M., Cao, Y., Hao, H., Xu, H., Aspuru-Guzik, A., et al. (2025). ORGANA: A robotic assistant for automated chemistry experimentation and characterization. *Matter* 8, 101897. <https://doi.org/10.1016/j.matt.2024.10.015>.
32. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-T., Rocktäschel, T., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. preprint at arXiv. <https://doi.org/10.48550/arXiv.2005.11401>.
33. Wang, B., Xu, C., Zhao, X., Ouyang, L., Wu, F., Zhao, Z., Xu, R., Liu, K., Qu, Y., Shang, F., et al. (2024). MinerU: An open-source solution for precise document content extraction. preprint at arXiv. <https://doi.org/10.48550/arXiv.2409.18839>.
34. Chen, J., Xiao, S., Zhang, P., Luo, K., Lian, D., and Liu, Z. (2024). M3-Embedding: multi-linguality, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. preprint at arXiv. <https://doi.org/10.48550/arXiv.2402.03216>.

35. Edge, D., Trinh, H., Cheng, N., Bradley, J., Chao, A., Mody, A., Truitt, S., Metropolitansky, D., Ness, R.O., and Larson, J. (2024). From local to global: A graph RAG approach to query-focused summarization. preprint at arXiv. <https://doi.org/10.48550/arXiv.2404.16130>. <https://arxiv.org/abs/2404.16130>.
36. Farquhar, S., Kossen, J., Kuhn, L., and Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature* 630, 625–630. <https://doi.org/10.1038/s41586-024-07421-0>.
37. Zeni, C., Pinsler, R., Zügner, D., Fowler, A., Horton, M., Fu, X., Wang, Z., Shysheya, A., Crabbé, J., Ueda, S., et al. (2025). A generative model for inorganic materials design. *Nature* 639, 624–632. <https://doi.org/10.1038/s41586-025-08628-5>.
38. Yang, H., Hu, C., Zhou, Y., Liu, X., Shi, Y., Li, J., Li, G., Chen, Z., Chen, S., Zeni, C., et al. (2024). Mattersim: A deep learning atomistic model across elements, temperatures and pressures. preprint at arXiv. <https://doi.org/10.48550/arXiv.2405.04967>.
39. Zhou, Y., Madaan, A., Liu, S., Yao, H., Hallinan, W., Yu, Q., Naik, A., and Yang, Y. (2022). Large language models are human-level prompt engineers. preprint at arXiv. <https://doi.org/10.48550/arXiv.2211.01910>.
40. Kojima, T., Gu, S.S., Reid, M., Matsuo, Y., and Iwasawa, Y. (2022). Large language models are zero-shot reasoners. preprint at arXiv. <https://doi.org/10.48550/arXiv.2205.11916>.
41. Li, G., Hammoud, H., Itani, H., Khizbullin, D., and Ghanem, B. (2023). CAMEL: Communicative agents for “mind” exploration of large language model society. preprint at arXiv. <https://doi.org/10.48550/arXiv.2303.17760>.
42. Wang, Z., Mao, S., Wu, W., Ge, T., Wei, F., and Ji, H. (2023). Unleashing the emergent cognitive synergy in large language models: A task-solving agent through multi-persona self-collaboration. preprint at arXiv. <https://doi.org/10.48550/arXiv.2307.05300>.
43. Zhang, J., Xu, X., Zhang, N., Liu, R., Hooi, B., and Deng, S. (2023). Exploring collaboration mechanisms for LLM agents: A social psychology view. preprint at arXiv. <https://doi.org/10.48550/arXiv.2310.02124>.
44. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. preprint at arXiv. <https://doi.org/10.48550/arXiv.2005.11401>.
45. Guu, K., Lee, K., Tung, Z., Pasupat, P., and Chang, M. (2020). REALM: Retrieval augmented language model pre-training. preprint at arXiv. <https://doi.org/10.48550/arXiv.2002.08909>.
46. Li, M., Zeng, Y., Qu, X., Jalalah, M., Alsareii, S.A., Li, C., Harraz, F.A., and Li, G. (2021). Biocatalytic CsPbX₃ perovskite nanocrystals: a self-reporting nanoprobe for metabolite analysis. *Small* 17, 2103255. <https://doi.org/10.1002/smll.202103255>.
47. Li, M., Tian, T., Zeng, Y., Zhu, S., Li, C., Yin, Y., and Li, G. (2021). One-step assay of pore-forming biotoxins based on biomimetic perovskite nanocrystals. *Sens. Actuators B* 338, 129839. <https://doi.org/10.1016/j.snb.2021.129839>.
48. Li, M., Wang, Y., Hu, H., Feng, Y., Zhu, S., Li, C., and Feng, N. (2022). A dual-readout sandwich immunoassay based on biocatalytic perovskite nanocrystals for detection of prostate specific antigen. *Biosens. Bioelectron.* 203, 113979. <https://doi.org/10.1016/j.bios.2022.113979>.
49. Yue, Y., Xu, H., Jiang, L., Zhao, X., and Deng, D. (2024). Introducing specific iodine ions in perovskite-based nanocomplex to cater for versatile biomedical imaging and tumor radiotherapy. *Adv. Healthcare Mater.* 13, 2302721. <https://doi.org/10.1002/adhm.202302721>.
50. Ghysels, A., Krämer, A., Venable, R.M., Teague, W.E., Lyman, E., Gawrisch, K., and Pastor, R.W. (2019). Permeability of membranes in the liquid ordered and liquid disordered phases. *Nature Commun* 10, 5616. <https://doi.org/10.1038/s41467-019-13432-7>.
51. Reber, P.J., and Squire, L.R. (2009). Intact visual discrimination learning in amnesic patients with hippocampal lesions. *Science* 325, 1514–1516. <https://doi.org/10.1126/science.1174621>.
52. de Meyer, F., and Smit, B. (2009). Effect of cholesterol on the structure of a phospholipid bilayer. *Proc. Natl. Acad. Sci. USA* 106, 3654–3658. <https://doi.org/10.1073/pnas.0809959106>.
53. Yang, Z., Xu, J., Zong, S., Xu, S., Zhu, D., Zhang, Y., Chen, C., Wang, C., Wang, Z., and Cui, Y. (2019). Lead halide perovskite nanocrystals-phospholipid micelles and their biological applications: multiplex cellular imaging and in vitro tumor targeting. *ACS Appl. Mater. Interfaces* 11, 47671–47679. <https://doi.org/10.1021/acsami.9b12924>.
54. Zhang, X., Xu, B., Zhang, J., Gao, Y., Zheng, Y., Wang, K., and Sun, X.W. (2016). All-inorganic perovskite nanocrystals for high-efficiency light emitting diodes: dual-phase CsPbBr₃-CsPb₂Br₅ composites. *Adv. Funct. Mater.* 26, 4595–4600. <https://doi.org/10.1002/adfm.201600958>.
55. Liu, Y., Li, F., Liu, Q., and Xia, Z. (2018). Synergetic effect of postsynthetic water treatment on the enhanced photoluminescence and stability of CsPbX₃ (X = Cl, Br, I) perovskite nanocrystals. *Chem. Mater.* 30, 6922–6929. <https://doi.org/10.1021/acs.chemmater.8b03330>.
56. Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E.P., et al. (2023). Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. preprint at arXiv. <https://doi.org/10.48550/arXiv.2306.05685>.
57. Chiang, W.-L., Zheng, L., Sheng, Y., Angelopoulos, A.N., Li, T., Li, D., Zhu, B., Zhang, H., Jordan, M.I., Gonzalez, J.E., et al. (2024). Chatbot arena: An open platform for evaluating LLMs by human preference. preprint at arXiv. <https://doi.org/10.48550/arXiv.2403.04132>.