

SciAgents: Automating Scientific Discovery Through Bioinspired Multi-Agent Intelligent Graph Reasoning

Alireza Ghafarollahi and Markus J. Buehler*

A key challenge in artificial intelligence (AI) is the creation of systems capable of autonomously advancing scientific understanding by exploring novel domains, identifying complex patterns, and uncovering previously unseen connections in vast scientific data. In this work, SciAgents, an approach that leverages three core concepts is presented: (1) large-scale ontological knowledge graphs to organize and interconnect diverse scientific concepts, (2) a suite of large language models (LLMs) and data retrieval tools, and (3) multi-agent systems with in-situ learning capabilities. Applied to biologically inspired materials, SciAgents reveals hidden interdisciplinary relationships that were previously considered unrelated, achieving a scale, precision, and exploratory power that surpasses human research methods. The framework autonomously generates and refines research hypotheses, elucidating underlying mechanisms, design principles, and unexpected material properties. By integrating these capabilities in a modular fashion, the system yields material discoveries, critiques and improves existing hypotheses, retrieves up-to-date data about existing research, and highlights strengths and limitations. This is achieved by harnessing a “swarm of intelligence” similar to biological systems, providing new avenues for discovery. How this model accelerates the development of advanced materials by unlocking Nature’s design principles, resulting in a new biocomposite with enhanced mechanical properties and improved sustainability through energy-efficient production is shown.

1. Introduction

One of the grand challenges in the evolving landscape of scientific discovery is finding ways to model, understand, and utilize information mined from diverse sources as a foundation for further research progress and new science discovery. Traditionally, this has been the domain of human researchers who review background knowledge, draft hypotheses, assess and test these hypotheses through various methods (in silico or in vitro), and refine them based on their findings. While these conventional approaches have led to breakthroughs throughout the history of science, they are constrained by the researcher’s ingenuity and background knowledge, potentially limiting discovery to the bounds of human imagination. Additionally, conventional human-driven methods are inadequate for exploring the vast amount of existing scientific data to extrapolate knowledge toward entirely novel ideas specially for multi-disciplinary areas like bio-inspired materials design where a common goal is to extract principles from Nature’s toolbox and bring it to bear towards engineering applications.

The emergence of AI technologies presents a potential promising solution by enabling the analysis and synthesis of large datasets beyond human capability, which could significantly accelerate discovery by uncovering patterns and connections that are not immediately obvious to human researchers.^[1-5] Therefore, there is great interest in developing AI systems that can not only explore and exploit existing knowledge to make significant scientific discoveries but also automate and replicate the broader research process, including acquiring relevant knowledge and data.^[6-10]

Large language models (LLMs), such as OpenAI’s GPT series,^[11] have demonstrated remarkable progress in diverse domains, driven by their robust capabilities.^[12-16] These foundational general-purpose AI models^[11,17-19] have been increasingly applied in scientific analysis, where they facilitate the generation of new ideas and hypotheses, offering solutions to some of the intrinsic limitations of conventional human-driven methods.^[20-27] Despite their successes, significant challenges persist regarding their ability to achieve the level of expertise possessed by domain specialists without extensive specialized training. Common

A. Ghafarollahi
 Laboratory for Atomistic and Molecular Mechanics (LAMM)
 Massachusetts Institute of Technology
 77 Massachusetts Ave., Cambridge, MA 02139, USA

M. J. Buehler
 Laboratory for Atomistic and Molecular Mechanics (LAMM)
 Center for Computational Science and Engineering
 Schwarzman College of Computing
 Massachusetts Institute of Technology
 77 Massachusetts Ave., Cambridge, MA 02139, USA
 E-mail: mbuehler@MIT.EDU

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/adma.202413523>

© 2024 The Author(s). Advanced Materials published by Wiley-VCH GmbH. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

DOI: [10.1002/adma.202413523](https://doi.org/10.1002/adma.202413523)

issues include their tendency to produce inaccurate responses when dealing with questions that fall outside their initial training scope, and broader concerns about accountability, explainability, and transparency. These problems underscore the potential risks associated with the generation of misleading or even harmful content, requiring us to think about strategies that increase their problem-solving and reasoning capabilities.

In response to these challenges, in-context learning emerges as a compelling strategy to enhance the performance of LLMs without the need for costly and time-intensive fine-tuning. This approach exploits the model's inherent ability to adapt its responses based on the context embedded within the prompt, which can be derived from a variety of sources. This capability enables LLMs to execute a wide array of tasks effectively.^[28–30] The potential to construct powerful generative AI models that integrate external knowledge to provide context and elicit more precise responses during generation is substantial.^[31] The central challenge is to develop robust mechanisms for the accurate retrieval and integration of relevant knowledge that enables LLMs to interpret and synthesize information pertinent to specific tasks, particularly in the realm of scientific discovery.

The construction of knowledge bases and the strategic retrieval of information from them are gaining traction as effective methods to enhance the generative capabilities of LLMs. Recent advancements in generative AI allow for the efficient mining of vast scientific datasets, transforming unstructured natural language into structured data such as comprehensive ontological knowledge graphs.^[6,32–35] These knowledge graphs not only provide a mechanistic breakdown of information but also offer an ontological framework that elucidates the interconnectedness of different concepts, delineated as nodes and edges within the graph.

While single-LLM-based agents can generate more accurate responses when enhanced with well-designed prompts and context, they often fall short for the complex demands of scientific discovery. Creating new scientific insights involves a series of steps, deep thinking, and the integration of diverse, sometimes conflicting information, making it a challenging task for a single agent. To overcome these limitations and fully leverage AI in automating scientific discovery, it's essential to employ a team of specialized agents. Multi-agent AI systems are known for their ability to tackle complex problems across different domains by pooling their capabilities.^[23,36–39] This collaborative approach allows the system to handle the intricacies of scientific discovery more effectively, potentially leading to breakthroughs that are difficult to achieve by single agents alone.

Building on these insights, our study introduces a method that synergizes the strengths of ontological knowledge graphs^[40,41] with the dynamic capabilities of LLM-based multi-agent systems, setting a robust foundation for enhancing graph reasoning and automating the scientific discovery process. Within this generative framework, the discovery workflow is systematically broken down into more manageable subtasks. Each agent in the system is assigned a distinct role, optimized through complex prompting strategies to ensure that every subtask is tackled with targeted expertise and precision. This strategic division of labor allows the AI system to proficiently manage the complexities of scientific research, fostering effective collaboration among agents. This collaboration is crucial for generating, refining, and crit-

ically evaluating new hypotheses against essential criteria like novelty and feasibility.

Central to our hypothesis generation is the utilization of a large ontological knowledge graph, focusing on biological materials, and developed from around 1000 scientific papers in this domain.^[6] We implemented a novel sampling strategy to extract relevant sub-graphs from this comprehensive knowledge graph, allowing us to identify and understand the key concepts and their interrelationships. This rich, contextually informed backdrop is crucial for guiding the agents in generating well-informed and innovative hypotheses. Such a method not only improves the accuracy of hypothesis generation but also ensures that these hypotheses are solidly rooted in a comprehensive knowledge framework. This structured approach promises to enhance the impact and relevance of scientific discoveries by ensuring they are well-informed and methodologically sound.

The plan of the paper is as follows. In Section 2, we discuss our proposed LLM-powered multi-agent system for automated scientific discovery, outlining its main components and constitutive agents. Two approaches are discussed and compared: One based on pre-programmed AI-AI interactions, and another one utilizing a fully automated framework in which a set of agents self-organize to solve problems. Several examples are provided to illustrate the different aspects of our approach, from path generation to research hypothesis generation and critique, demonstrating the system's potential to explore novel scientific concepts and produce innovative ideas by synthesizing an iterative prompting strategy during which multiple LLMs work together. Section 3 then presents the key findings and discussing the implications of our multi-agent system for future research in scientific discovery.

2. Results and Discussion

LLMs have demonstrated a relatively high level of proficiency across a wide range of tasks, including question answering, hypothesis development, summarizing and contrasting ideas, processing complex information, executing tasks, and even writing code. However, conventional inference strategies often fail to produce sophisticated reasoning and detail in the generated data. By using a set of interacting models, and assigning distinct roles to LLM-based agents, effective multi-agent AI systems can be constructed. When combined with carefully crafted prompts and in-context learning from graph representation of data, these systems are capable of generating scientific ideas and hypotheses. We now present results from a several experiments we conducted with our proposed framework (details about implementation, see Experimental Section).

2.1. Multi-Agent System for Graph Reasoning and Scientific Discovery

Figure 1 illustrates the outline of our proposed multi-agent model designed to automate the scientific discovery process based on the key concepts and relationships retrieved from a comprehensive knowledge graph developed from scientific papers (Figure 1a). This figure further showcases two distinct strategies

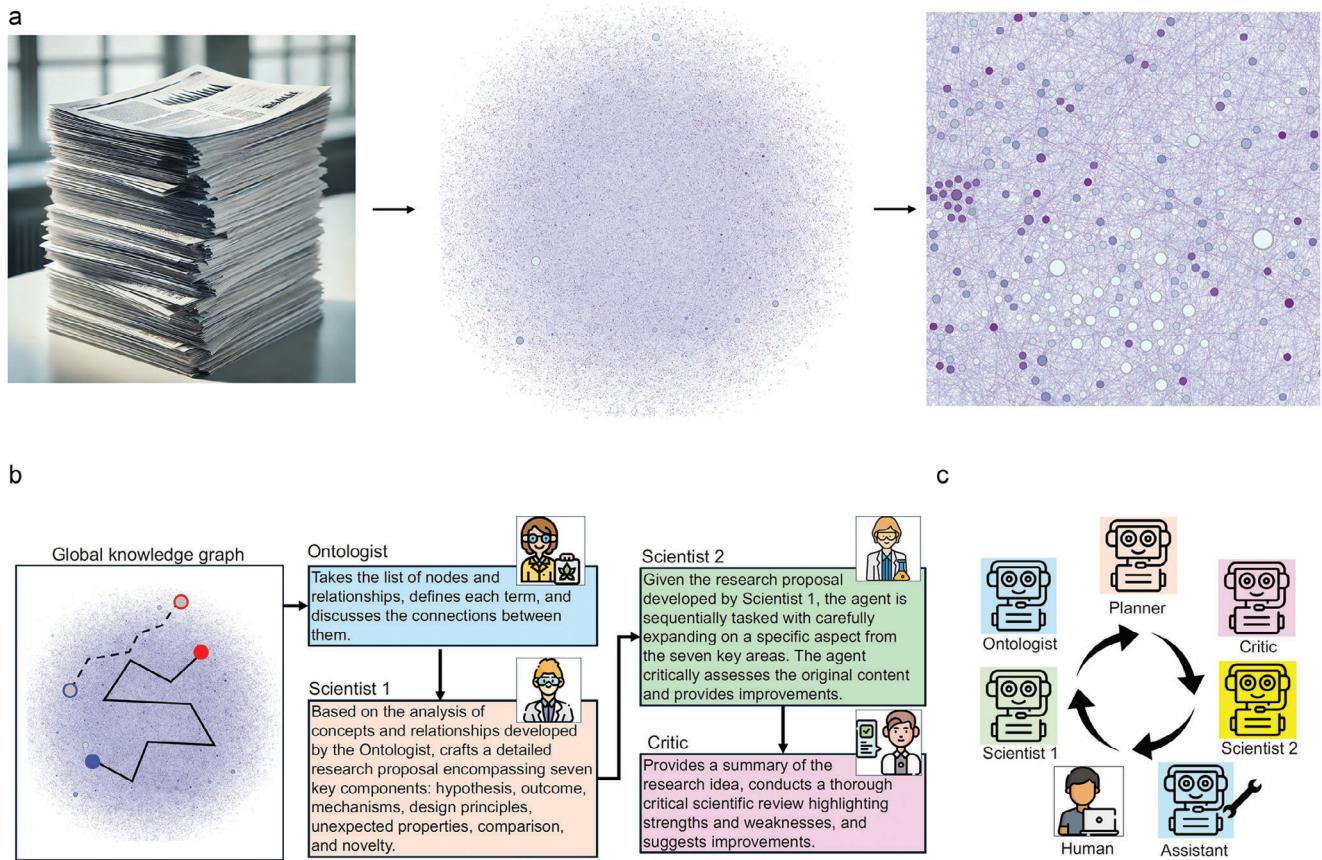


Figure 1. Overview of the multi-agent graph-reasoning system developed here. Panel a, overview of graph construction, as reported in Ref. [6]. The visual shows the progression from scientific papers as data source to graph construction, with the image on the right showing a zoomed-in view of the graph. Panels b and c: Two distinct approaches are presented: In panel b, A multi-agent system based on pre-programmed sequence of interactions between agents, ensuring consistency and reliability, and in panel c, a fully automated, flexible multi-agent framework that adapts dynamically to the evolving research context. Both systems leverage a sampled path within a global knowledge graph as context to guide the research idea generation process. Each agent plays a specialized role: The Ontologist defines key concepts and relationships, Scientist 1 crafts a detailed research proposal, Scientist 2 expands and refines the proposal, and the Critic agent conducts a thorough review and suggests improvements. The Planner in the second approach develops a detailed plan and the assistant is instructed to check the novelty of the generated research hypotheses. This collaborative framework enables the generation of innovative and well-rounded scientific hypotheses that extend beyond conventional human-driven methods.

deployed in this study for generating novel scientific hypotheses, both of which harness the collective intelligence of a team of agents. These strategies integrate the specialized capabilities of each agent, systematically exploring uncharted research territories to produce innovative and high-impact scientific hypotheses. The full description of the agents incorporated in SciAgents is listed in Figures S1–S4 (Supporting Information).

The key difference between these approaches lies in the nature of the interaction between the agents. In the first approach (Figure 1b), the interactions between agents are pre-programmed and follow a predefined sequence of tasks that ensure consistency and reliability in hypothesis generation. In contrast, the second approach features fully automated agent interactions without any predetermined order of how interactions between agents unfold, providing a more flexible and adaptive framework that can dynamically respond to the evolving context of the research process. This second strategy (Figure 1c) also incorporates human-in-the-loop interactions, enabling human intervention at various stages of research development. Such interventions allow for ex-

pert feedback, refinement of hypotheses, or strategic guidance, specification about certain materials, types or features, ultimately enhancing the quality and relevance of the generated scientific ideas. Moreover, the second approach provides a more robust framework where additional tools could be readily incorporated. For instance, we have empowered our automated multi-agent model with the Semantic Scholar API as a tool that provides it with an ability to check the novelty of the generated hypothesis against the existing literature.

Figure 2 shows an overview of the entire process from initial keyword selection to the final document. We employ a hierarchical expansion strategy where answers are successively refined and improved, enriched with retrieved data, critiqued and amended by identification or critical modeling, simulation and experimental tasks and adversarial prompting. The process begins with initial keyword identification or random exploration within a graph, followed by path sampling to create a subgraph of relevant concepts and relationships. This subgraph forms the basis for generating structured output in JSON following a specific

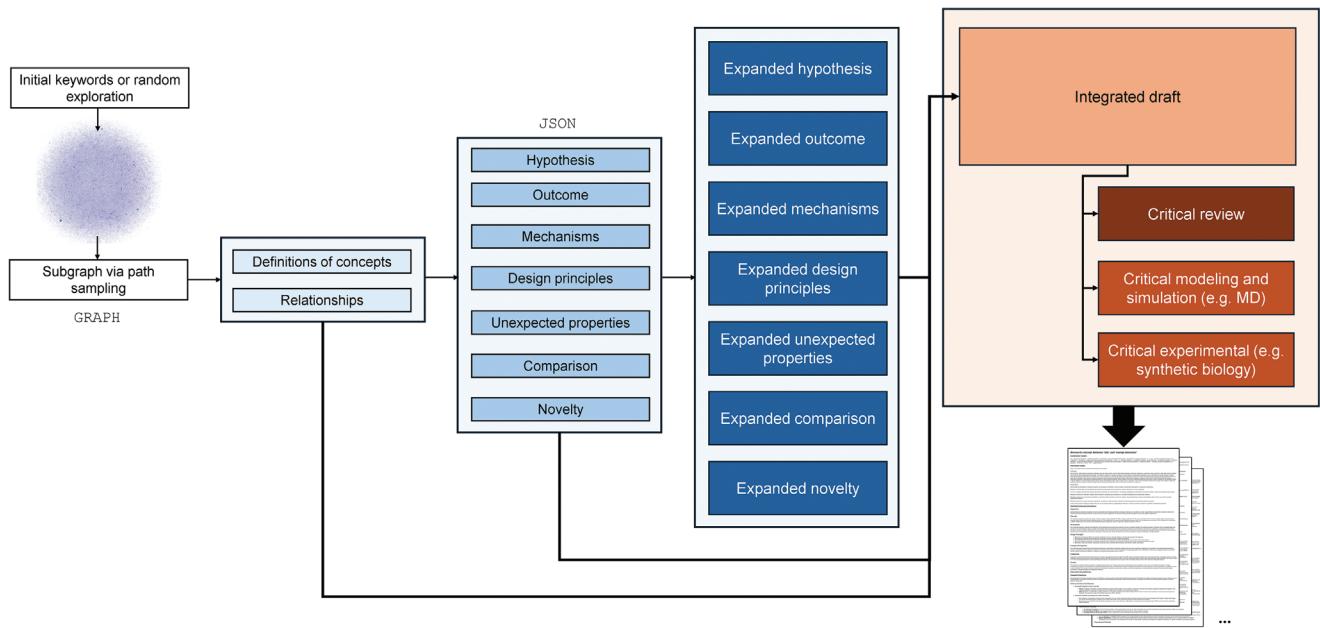


Figure 2. Overview of the entire process from initial keyword selection to the final document, following a hierarchical expansion strategy where answers are successively refined and improved, enriched with retrieved data, critiqued and amended by identification or critical modeling, simulation and experimental tasks. The process begins with initial keyword identification or random exploration within a graph, followed by path sampling to create a subgraph of relevant concepts and relationships (see, Figure S1, Supporting Information, for an illustration of how the path can be sampled). This subgraph forms the basis for generating structured output in JSON, including the hypothesis, outcome, mechanisms, design principles, unexpected properties, comparison, and novelty. Each component is subsequently expanded on with individual prompting, to yield significant amount of additional detail, forming a comprehensive draft. This draft then undergoes a critical review process, including amendments for modeling and simulation priorities (e.g., molecular dynamics) and experimental priorities (e.g., synthetic biology). The final integrated draft, along with critical analyses, results in a document that guides further scientific inquiry.

set of aspects that the model is tasked to develop. These include the hypothesis, outcome, mechanisms, design principles, unexpected properties, comparison, and novelty. The details of each component can be found in Figure S5 (Supporting Information). For instance, the “outcome” describes the expected findings or impact of the research, the “mechanisms” field discusses predicted chemical, biological, or physical interactions, while “unexpected properties” anticipates emergent behaviors from novel combinations of concepts in the graph. Each component is subsequently expanded on with individual prompting, to yield significant amount of additional detail, forming a comprehensive draft. This draft then undergoes a critical review process, including amendments for modeling and simulation priorities (e.g., molecular dynamics) and experimental priorities (e.g., synthetic biology). The final integrated draft, along with critical analyses, results in a document that can guide further scientific inquiry.

In the following, we explore the primary components of our multi-agent strategy. For better clarity and understanding, each section is accompanied by practical examples from a sample hypothesis. This hypothesis was generated using “silk” and “energy-intensive” as the starting nodes. The outcomes of this experiment are presented in Figure 3. For a more detailed illustration, see the Supplementary Information.

1- Path Generation: At the core of our model is an expansive knowledge graph, first introduced in Ref. [6], that encompasses the fields of bio-inspired materials and mechanics. This knowledge graph integrates a variety of concepts and knowledge domains, enabling the exploration of hypotheses that once

seemed disconnected. To augment the capabilities of our underlying large language model (LLM), we provide it with a subgraph derived from this more extensive knowledge graph. This sub-graph depicts a pathway that connects two crucial concepts or nodes within the comprehensive graph. The construction of this path is crucial; Unlike in earlier work^[6] where the shortest path was utilized, our study employs a random path approach. As illustrated in Figure 4, the random approach infuses the path with a richer array of concepts and relationships, enabling our agents to explore a broader spectrum of domains, as opposed to the shortest path where only a few concepts are included. This expanded exploration not only enhances the depth and breadth of insights gained but also fosters the novelty of the hypotheses generated. Initially, the two concepts can be either specified by the user or selected randomly by the model from the knowledge graph. For instance, the example below demonstrates the path generated by the model between the concepts “silk” and “energy-intensive”. Figure 8 shows additional knowledge graphs derived from random sampling for randomly chosen concepts to provide additional examples. We refer the reader to Figure S1 (Supporting Information) for a visualization of how path sampling can be conducted between two predetermined nodes, or randomly selected pairs of nodes.

silk → provides → biocompatibility → possess → biological materials → has → multifunctionality → include →

1-521405_0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/adma.202413523. Wiley Online Library is available on [12/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License.



Figure 3. Results from our multi-agent model, illustrating a novel research hypothesis based on a knowledge graph connecting the keywords “silk” and “energy-intensive,” as an example. This visual overview shows that the system produces detailed, well-organized documentation of research development with multiple pages and detailed text (the example shown here includes 8100 words). Details of the results are presented in the main text and other figures, and full conversations generated by the SciAgents model are included as Supporting Information.

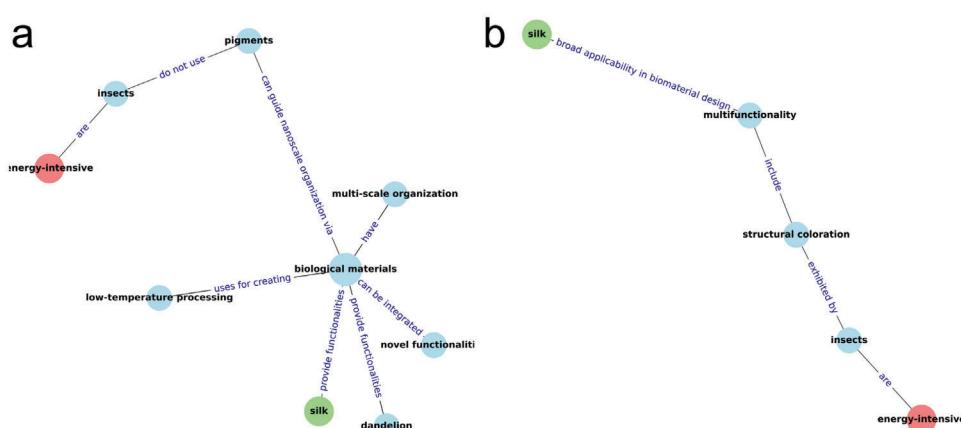


Figure 4. The knowledge graphs connecting the keywords “silk” and “energy-intensive” extracted from the global graph using a) random path and b) the shortest path between the concepts. The difference between nodes and edges sampled in the two approaches is apparent, where enhanced sampling invokes a host of additional concepts that will be incorporated into research development. The richer substrate that forms the basis for agentic reasoning yields more sophisticated research concepts. Agentic reasoning carefully assesses the ideas and negotiate, via adversarial interactions between the agents, a sound prediction and carefully delineated research ideas [36,38,39,42–44].

self-cleaning -> include -> multifunctionality -> broad applicability in biomaterial design -> silk -> possess -> biopolymers -> possess -> silk -> is -> fibroin -> is -> silk -> broad applicability in biomaterial design -> multifunctionality -> include -> structural coloration -> exhibited by -> insects -> are -> **energy-intensive**

- ****Structural coloration – exhibited by – insects**:** Insects, such as those that produce silk, often exhibit structural coloration.
- ****Insects – are – energy-intensive**:** The processes involving insects, including silk production, can be energy-intensive.

The generated path provides an analytical representation of various concepts and their interconnections, which were previously unrelated. By delineating these relationships, the model gains the ability to perceive and analyze connections between concepts that have not been explicitly linked before. This innovative mapping approach enables the model to extrapolate and generate ideas that are both novel and potentially transformative, paving the way for breakthroughs in understanding and application.

2- Deep Insights with LLM-Based Analysis: Utilizing our LLM-powered ontologist agent, we move deeper into the intricacies of the relationships that have been mapped out in the earlier path generation stage. By examining the connections and nuances among the identified concepts, the agent helps transition from static knowledge retrieval to dynamic knowledge generation. This crucial shift is what enables the model to identify gaps in existing research and propose new angles of inquiry, thereby laying the groundwork for novel ideas and hypotheses. In this context, the role of the ontologist agent is instrumental. It applies advanced reasoning and inference techniques to synthesize and interpret the complex web of data. This capability allows it to extract significant insights that might not be obvious at first glance, offering a richer, more detailed understanding of the relationships.

As an example, here are some of the insights into the relationships as identified by the ontologist for the aforementioned path, demonstrating how our approach not only maps out but also meaningfully interprets complex datasets to generate scientific hypotheses.

- ****Silk – possess – biopolymers**:** Silk is a type of biopolymer, a natural polymer produced by living organisms.
- ****Biopolymers – possess – silk**:** This reiterates that silk is a biopolymer.
- ****Silk – can be spun into – membranes**:** Silk can be processed and spun into thin layers or sheets known as membranes.
- ****Membranes – can be spun into – silk**:** This indicates the reverse process, where membranes can be derived from silk.
- ****Silk – broad applicability in biomaterial design – multifunctionality**:** Silk's multifunctional properties make it highly applicable in the design of biomaterials.
- ****High adhesion – include – multifunctionality**:** This indicates that high adhesion is part of the multifunctional nature of silk.

The results demonstrate that the model has developed a reasonably refined understanding of relationships between seemingly unrelated concepts. This capability enables the model to support reasoning in scientific research and propose new research hypotheses, which will be further explored in the subsequent stage.

3- Research Hypothesis Generation and Expansion: This stage is where the effects of our multi-agent system emerges. The scientist agent harnesses the extensive knowledge parsed from the knowledge graph and further refined by the ontologist to propose novel research ideas. Through complex prompting, as shown in Figure 5, the agent is assigned specific roles and is tasked with synthesizing a novel research proposal that integrates all key concepts from the knowledge graph. The designated agent, Scientist_1, is configured to deliver a detailed hypothesis that is both innovative and logically grounded, aiming to advance the understanding or application of the provided concepts. The agent creates a proposal that carefully addresses the following seven key aspects: hypothesis, outcome, mechanisms, design principles, unexpected properties, comparison, and novelty. This approach ensures a thorough exploration and evaluation of the new scientific idea, allowing for a detailed assessment of its feasibility, potential impact, and areas of innovation.

The proficiency of the Scientist_1 LLM agent in generating novel research hypotheses is demonstrated in Figure 3. The concept involves integrating silk with dandelion-based pigments to create biomaterials with enhanced optical and mechanical properties. The proposed enhancement in mechanical properties stems from a hierarchical organization of silk combined with the reinforcing effects of the pigments. According to the model, this proposed composite material could exhibit significantly improved mechanical strength, reaching up to 1.5 GPa compared to traditional silk materials, which range from 0.5 to 1.0 GPa. Additionally, the use of low-temperature processing and dandelion pigments is projected to reduce energy consumption by approximately 30%. This example underscores the potential of translating knowledge graphs into unprecedented material designs, facilitating a seamless transition from theoretical data to practical applications in materials science.

The research idea proposed by Scientist_1 provides a foundational abstract for a more detailed research proposal that is developed through subsequent agentic interactions. To enhance and deepen this initial concept, Scientist_2 is tasked with rigorously expanding upon and critically assessing the idea's various components. This agent is specifically instructed to integrate, whenever possible, quantitative scientific information such as chemical formulas, numerical values, protein sequences, and processing conditions, significantly enriching the proposal's scientific depth and accuracy. Additionally, Scientist_2 is directed to comment on specific modeling and simulation techniques tailored

You are a sophisticated scientist trained in scientific research and innovation.

...
Analyze the graph deeply and carefully, then craft a detailed research hypothesis that investigates a likely groundbreaking aspect that incorporates EACH of these concepts. Consider the implications of your hypothesis and predict the outcome or behavior that might result from this line of investigation. Your creativity in linking these concepts to address unsolved problems or propose new, unexplored areas of study, emergent or unexpected behaviors, will be highly valued. Be as quantitative as possible and include details such as numbers, sequences, or chemical formulas. Please structure your response in JSON format, with SEVEN keys:

"hypothesis" clearly delineates the hypothesis at the basis for the proposed research question.
"outcome" describes the expected findings or impact of the research. Be quantitative and include numbers, material properties, sequences, or chemical formula.
"mechanisms" provides details about anticipated chemical, biological or physical behaviors. Be as specific as possible, across all scales from molecular to macroscale.
"design_principles" should list out detailed design principles, focused on novel concepts and include a high level of detail. Be creative and give this a lot of thought, and be exhaustive in your response.
"unexpected_properties" should predict unexpected properties of the new material or system. Include specific predictions, and explain the rationale behind these clearly using logic and reasoning. Think carefully.
"comparison" should provide a detailed comparison with other materials, technologies or scientific concepts. Be detailed and quantitative.
"novelty" should discuss novel aspects of the proposed idea, specifically highlighting how this advances over existing knowledge and technology.

Ensure your scientific hypothesis is both innovative and grounded in logical reasoning, capable of advancing our understanding or application of the concepts provided.

Figure 5. The profile of the Scientist_1 LLM agent implemented in the first proposed multi-agent approach for automated scientific discovery. The AI agent utilizes the definitions of concepts and relationships between them in the knowledge graph provided by the Ontologist to generate a novel research hypothesis.

to the project's needs, such as simulations for material behavior analysis or experimental methods. This thorough review and enhancement process, including clear rationale and step-by-step reasoning, ensures that the research proposal is robust, well-grounded, and ready for further development. This systematic approach not only solidifies the scientific underpinnings of the proposal but also prepares it for successful implementation and future exploration.

The expanded research idea provided by Scientist_2 is showcased in the Supplementary Information, revealing a thorough rationale and sequential reasoning for various aspects of the research proposal. Here are selected key points to exemplify the model's contributions:

- The model suggests using Molecular Dynamics (MD) Simulations to explore interactions at the molecular level. Specifically, it proposes employing software like GROMACS or AMBER to model how silk fibroin interacts with dandelion pigments, aiming to understand the self-assembly processes and predict the resulting microstructures.
- For potential applications of the new composite material, the model identifies its suitability for bio-inspired adhesives. It highlights how the dynamic interactions between silk proteins and pigments may impart self-healing properties, mak-

ing these materials ideal for adhesives that can repair themselves after damage.

- Regarding the mechanisms that contribute to enhanced material properties, the model points out the reinforcing effect of the pigments. It suggests that these pigments could improve the tensile strength and toughness of the composite material, with plans to conduct mechanical testing, including tensile and nanoindentation tests, to quantify these properties.
- A detailed comparison with existing materials is also provided by the model as summarized in **Table 1**. It notes that traditional silk materials typically exhibit tensile strengths ranging from 0.5 to 1.0 GPa, whereas the proposed composite material aims to achieve up to 1.5 GPa. This enhancement is attributed to the hierarchical organization of silk proteins and the reinforcing effect of dandelion-derived pigments. Further, it details how silk fibroin's molecular structure, with repetitive sequences of glycine and alanine forming β -sheet crystallites, contributes to its mechanical properties. The integration of dandelion pigments, possibly including bioactive compounds such as taraxasterol and luteolin, is expected to further enhance these properties through intermolecular interactions and cross-linking, providing a synergistic effect at multiple scales.
- As summarized in **Table 2**, the model proposes the following design principles: It utilizes the natural multi-scale

Table 1. Comparison of traditional silk materials versus proposed composite material, as predicted by our model.

Feature	Traditional Silk Materials	Proposed Composite Material	Details
Mechanical Strength	Tensile strength: 0.5–1.0 GPa.	Aiming for tensile strength up to 1.5 GPa.	Enhanced by hierarchical organization of silk fibroin (composed of Gly-Ala repeats forming β -sheet crystallites) and dandelion-derived pigments like taraxasterol ($C_{30}H_{50}O$) and luteolin ($C_{15}H_{10}O_6$).
Structural Colors	Requires synthetic dyes for color.	Utilizes dandelion-derived pigments for structural colors.	The pigments will self-assemble into nanoscale structures, such as photonic crystals or Bragg stacks, which can reflect specific wavelengths of light. The concentration and distribution of pigments will be optimized to achieve the desired optical properties
Energy Efficiency	Energy-intensive, high-temperature processing (boiling in Na_2CO_3 solution at $\sim 100^\circ C$).	Low-temperature processing below $50^\circ C$, reducing energy consumption by $\sim 30\%$.	The energy savings can be quantified by comparing the energy required for traditional silk degumming (typically involving boiling in alkaline solutions) with the energy required for the proposed low-temperature extraction and processing methods.

organization of silk fibroin to guide the self-assembly of dandelion pigments, leveraging hierarchical structuring from the nano to the macro scale. This organization is critical for achieving both the desired mechanical strength and vibrant structural coloration. The model emphasizes the need to control pigment concentration and distribution to ensure optimal optical properties, such as precise reflectance peaks, while maintaining flexibility and tensile strength. Moreover, it advocates for low-temperature processing to preserve the biocompatibility and structural integrity of silk proteins, ensuring

energy-efficient production methods. These principles collectively contribute to the creation of an advanced bio-inspired material with enhanced mechanical and optical functionalities.

- The model predicts unexpected properties including self-healing properties due to the dynamic nature of the silk-pigment interactions, stimuli-responsive structural colors as the structural colors could change in response to environmental stimuli, and additional functionalities such as UV protection and antimicrobial properties due to the bioactive

Table 2. Summary of design principles for energy-efficient, structurally colored silk composites.

Stage	Process Details	Methods
Low-Temperature Processing for Silk	Maintain temperatures below $50^\circ C$ during silk protein extraction and processing. Use aqueous solutions with a mild pH (6.5–7.5) to avoid denaturation. Monitor thermal stability with DSC	Differential scanning calorimetry (DSC) and circular dichroism (CD) spectroscopy to monitor the thermal stability of silk proteins.
Self-Assembly of Dandelion Pigments	Utilize the alignment of silk nanofibrils and microfibrils to guide the organization of dandelion-derived pigments. Predict interactions using MD simulations. Visualize with AFM and SEM.	Molecular dynamics (MD) simulations to predict the interaction energies between silk proteins and dandelion-derived pigments. Atomic force microscopy (AFM) and scanning electron microscopy (SEM) to visualize the hierarchical organization of pigments within the silk.
Pigment Concentration Optimization	Control pigment concentration within 0.1–1.0 wt% to achieve desired optical properties. Use FDTD simulations to model light interaction. Verify reflectance peaks with UV-Vis spectroscopy.	Use UV-Vis spectroscopy to analyze the optical properties and confirm the presence of desired reflectance peaks.
Hierarchical Structuring for Strength	Align and cross-link silk nanofibrils and microfibrils. Introduce cross-linking agent genipin ($C_{11}H_{14}O_5$). Analyze mechanical properties with FEA and DMA. Target tensile strength of 1.5 GPa.	Use FEA to simulate the mechanical behavior of the composite under different loading conditions. Use dynamic mechanical analysis (DMA) to study the viscoelastic properties and ensure a balance between strength and flexibility.
Energy-Efficient Production	Implement enzymatic extraction methods for silk proteins and pigments at low temperatures. Monitor energy usage with calorimetry. Evaluate sustainability with LCA. Aim for 30% energy reduction.	Use life cycle assessment (LCA) to evaluate the environmental impact and energy efficiency of the production process.

Table 3. Unexpected properties predicted for the silk-pigment composite material.

Self-Healing Properties	Mechanism	Details
Self-Healing Properties	Silk proteins (fibroin) re-form hydrogen bonds and β -sheet structures. Bioactive compounds in dandelion-derived pigments (e.g., taraxasterol) enhance self-healing through hydrogen bonding and hydrophobic interactions.	Recovery of mechanical strength can reach up to 80% within 24 hours at ambient conditions after damage. Self-healing efficiency is measured by the recovery of mechanical strength.
Stimuli-Responsive Structural Colors	The hygroscopic nature of silk and the responsive behavior of dandelion pigments cause swelling or contraction, altering the spacing between pigment nanoparticles and shifting the reflectance peak in response to humidity and temperature changes.	The reflectance peak shifts by 10-50 nm for a 10% change in relative humidity. This is measured using spectrophotometry and modeled using finite element analysis (FEA).
Additional Functionalities	Dandelion pigments introduce UV protection (via luteolin and caffeic acid) and antimicrobial properties (via taraxacin and sesquiterpene lactones), which absorb UV light and inhibit microbial growth.	UV protection efficiency exceeds 90%, and antimicrobial properties exhibit inhibition zones of 10-15 mm against <i>E. coli</i> and <i>S. aureus</i> . Measured through UV-Vis spectroscopy and antimicrobial assays.

compounds present in dandelions. Scientist 2 provides more details regarding the mechanisms underlying these properties as tabulated in **Table 3**.

At the final stage of our research development process is the Critic agent, responsible for thoroughly reviewing the research proposal, summarizing its key points, and recommending improvements. This agent delivers a comprehensive scientific critique, highlighting both the strengths and weaknesses of the research idea while suggesting areas for refinement. Additionally, the Critic agent is tasked to identify the most impactful scientific question that can be addressed through molecular modeling (e.g., molecular dynamics) and experimentation (e.g., synthetic biology), and to outline the critical steps for setting up and conducting these molecular and experimental priorities.

For our model example involving the silk-pigment composite material, the full response from the Critic is detailed in the Supporting Information. It provides a comprehensive evaluation of the proposed research methodology and its potential impact. The critic agent commends the integration of silk-derived biological materials with dandelion-based pigments for creating energy-efficient, structurally colored biomaterials, noting the project's interdisciplinary approach and innovative use of natural hierarchical structures to enhance mechanical and optical properties. The agent also recognizes the robustness added by the combined use of modeling techniques and experimental methods.

Moreover, the critic identifies areas needing improvement, including challenges with nanoscale integration, scalability, envi-

ronmental impacts of solvent use, and a lack of quantitative data. Concerns about the long-term stability of the material under real-world conditions are also raised. To address these issues, the critic suggests conducting pilot studies for process validation, exploring green chemistry for pigment extraction, developing detailed scalability plans, and performing rigorous analyses of energy consumption and material durability. These suggestions aim to refine the research direction, making the hypotheses generated by the AI system not only innovative but also practical, thereby enhancing the potential for significant scientific advancements.

Lastly, the critic proposes the most impactful scientific questions related to molecular modeling, simulation, and synthetic biology experiments as shown in **Figure 6**.

For each aspect, the critic agent provides detailed responses, outlining the key steps for setting up and conducting atomistic simulations and experimental work. To perform the molecular modeling and simulation, the critic describes the process of simulating the interaction and self-assembly of silk fibroin and dandelion-derived pigments using molecular dynamics (MD) simulations. This begins by defining the molecular structures of silk fibroin, rich in glycine and alanine, and key pigments like luteolin and taraxanthin, sourced from protein and chemical databases. Appropriate force fields, such as CHARMM or AMBER, are selected, with parameters defined using tools like CGenFF. The system is then prepared by placing the molecules in a solvated environment, adding ions for neutralization, and using VMD or GROMACS for setup. After energy minimization and equilibration under constant temperature and pressure, MD

Critic

How do the molecular interactions between silk fibroin and dandelion-derived pigments influence the self-assembly process and the resulting nanoscale structures that contribute to the composite material's mechanical and optical properties?

How can synthetic biology be used to engineer silk-producing organisms to incorporate dandelion-derived pigments directly into silk fibers during production, thereby creating structurally colored silk with enhanced mechanical properties and energy-efficient production?

Figure 6. Most impactful questions raised by the critic agent for the generated research hypothesis on integrating silk with dandelion-based pigments to create biomaterials with enhanced optical and mechanical properties.

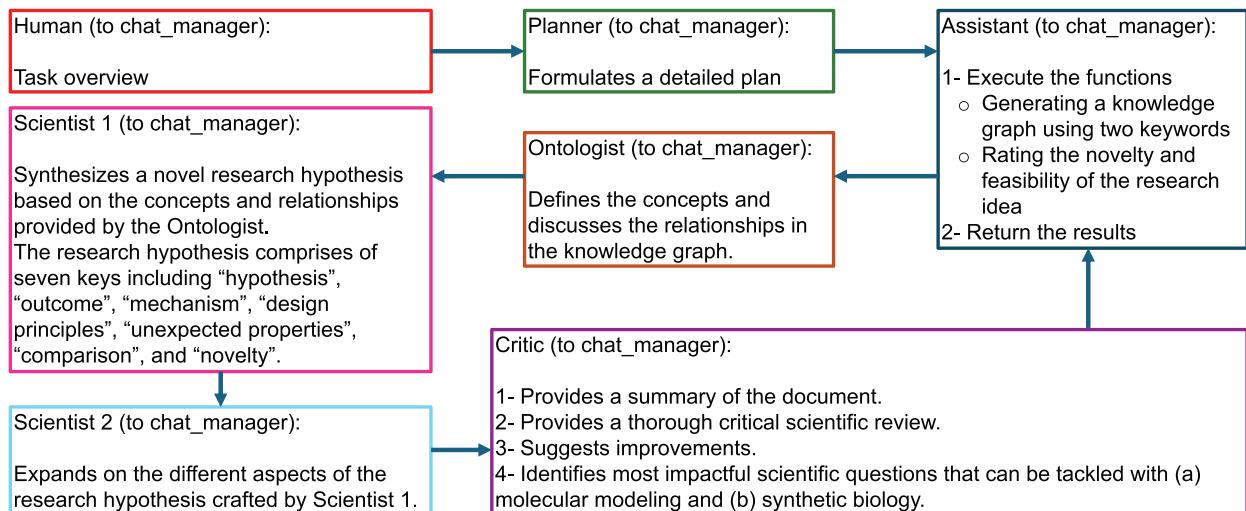


Figure 7. Flowchart illustrating the dynamic interactions as developed autonomously by the multi-agent team members, coordinated by the group chat manager, to generate research hypotheses through graph reasoning. The manager selects the working agents to collaborate based on the current chat context, fostering cooperation and enabling mutual adjustments to solve the problem.

simulations are run for 100–500 ns, using periodic boundary conditions. Post-simulation analysis includes calculating interaction energies, identifying binding sites, and performing cluster analysis of self-assembled structures, focusing on nanoscale formations like β -sheets in silk fibroin using tools like PyMOL, Chimera, and GROMACS.

We find that the critic agent plays a crucial role in guiding these efforts by posing probing scientific questions that challenge the assumptions and focus of the research, ensuring that the simulations and experiments target key mechanisms and outcomes. By doing so, the critic not only helps refine the research direction but also enhances the potential for discovering novel biomaterials with optimized mechanical and optical properties. This iterative feedback loop between hypothesis generation and critical evaluation strengthens the overall scientific process.

2.2. Autonomous Agentic Modeling

The experiments so far were conducted using the non-automated multi-agent system (see Figure 1), whereas the second approach described in this section uses an automated way to generate a research hypothesis from a knowledge graph that facilitates dynamic interactions.

The automated multi-agent system consists of a team of AI agents, each powered by a state-of-the-art general purpose large language model from the GPT-4 family,^[11] accessed via the OpenAI API.^[45] Each agent has a specific role and focus in the system which is described by a unique profile. Our team of agents with the following entities collaborate in a dynamic environment to create a research proposal:

- “Human”: human user that poses the task and can intervene at various stages of the problem solving process.
- “Planner”: suggests a detailed plan to solve the task.
- “Ontologist”: who is responsible to define the relationships and concepts within the knowledge graph.

- “Scientist 1”: crafts the initial draft of a detailed research hypothesis with seven key items based on the definitions provided by Ontologist.
- “Scientist 2”: who expands and refines the different key aspects of the research proposal created by Scientist 1.
- “Critic”: conducts a thorough review and suggests improvements.
- “Assistant”: has access to external tools including a function to generate a knowledge path from two keywords and a function to assess the novelty and feasibility of the research idea.
- “Group chat manager”: chooses the next speaker based on the context and agent profiles and broadcasts the message to the whole group.

Despite the varied dynamics in agentic AI–AI interactions, the overall pipeline of the two proposed agent-based systems to generate research hypotheses from concepts and relationships derived from a knowledge graph is similar. As illustrated in Figure 7 the automated multi-agent collaboration starts with a plan from the planner detailing the steps required to accomplish the task posed by the human which involves creating a research hypothesis from given keywords or randomly selected by the model. Next, the assistant agent calls the appropriate function to establish a pathway which serves as the foundational knowledge graph for subsequent analysis. The ontologist agent then discusses definitions and relationships. This sets the stage for scientist_1 to generate a research idea, which is then expanded by scientist_2. The sequence concludes with a summary, critical review, and suggestions for improvement by the critic agent. Finally, the assistant agent executes another tool to analyze and score the novelty and feasibility of the proposed research idea.

Despite the similarity in the steps followed by the agents in each approach, the results show that while the generated hypotheses share overall concepts and methodologies, they differ in the details. For example, in the analysis of the research hypothesis highlighted earlier, both models emphasize integrating silk

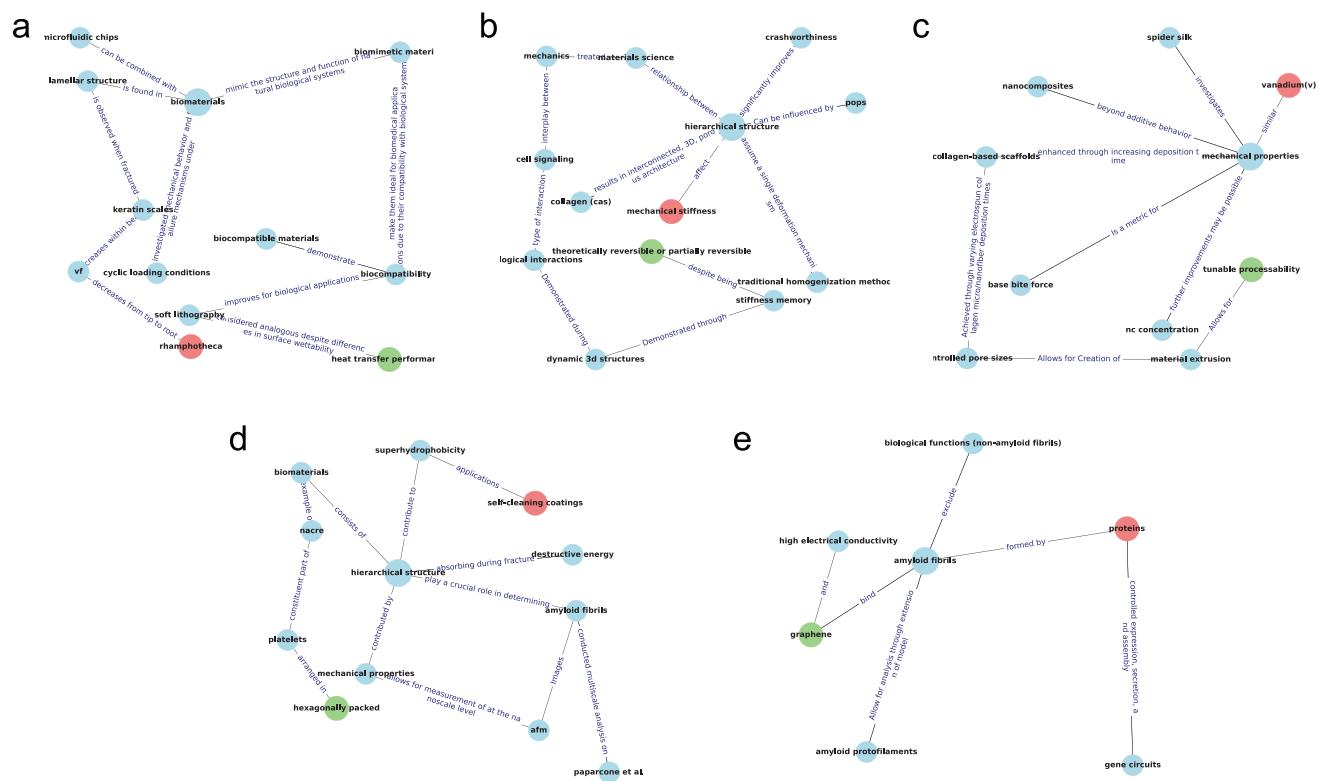


Figure 8. Knowledge graphs derived from random sampling for randomly chosen concepts from the global knowledge graph. Panel a: “heat transfer performance” connecting “rhamphotheca,” panel b: “theoretically reversible or partially reversible” connecting “mechanical stiffness”, panel c: “tunable processability” connecting “vanadium(v),” and panel d: “hexagonally packed” connecting “self-cleaning coating,” and panel e: “graphene” connecting “proteins”.

with dandelion pigments, but they differ in specifics such as their scope of application and the depth of technical aspects regarding material fabrication and potential uses. For comparison, the full document created by the automated multi-agent model using the same knowledge graph between “silk” and “energy-intensive” is provided in Section S2 (Supporting Information).

The difference stems from the subtle differences in how the data is propagated between the agents in the two approaches. In the first approach, during the generation process, the agents receive only a filtered subset of information from previous interactions (see Section 4.3 for more details). In contrast, the second approach allows agents to share memory, giving them access to all the content generated in previous interactions. This means they operate with full visibility of the history of their collaboration. Another difference between the two models is that the second approach benefits from the integration of a tool that assesses the novelty of the proposed research ideas against current literature, using Semantic Scholar API. This feature enables us to effectively measure the novelty of the research and proactively eliminate any ideas that are too similar to existing work.

To demonstrate the efficacy of the automated multi-agent model in generating novel research ideas and evaluating their novelty and feasibility, we conducted five experiments, tasking the automated multi-agent model with constructing research

ideas. We summarized these hypotheses in **Table 4**, which includes details about each research idea, the proposed hypotheses, expected outcomes, and assessments of novelty and feasibility. These research ideas are generated based on randomly selected concepts from the knowledge graph. **Figure 8** displays the generated knowledge graphs, showcasing a diverse array of concepts and relationships. Some nodes like “biomaterials,” “hierarchical structure,” and “mechanical properties” show high node degree and serve as central hubs, indicating their pivotal roles in interconnecting various scientific disciplines within the graph. The results highlight the diversity of the research hypotheses, which stems from both the random selection of endpoint nodes and the paths between them. Moreover, the results showcase varying levels of novelty and feasibility, as assessed against current literature, underscoring the critical role of comparing with existing knowledge. The process of exploring a variety of paths, scoring the results, and identifying the most promising directions could easily be scaled over thousands of iterations, yielding a very large ideation database.

Below, we provide additional details on the various aspects of the research hypotheses for a selected sample. The complete documents for the five hypotheses can be found in Sections S3–S7 (Supporting Information).

An example of a research hypothesis generated with the knowledge graph depicted in Figure 8a is provided in Section S3

Table 4. Examples of research ideas generated by SciAgents using automated approach featuring underlying hypothesis, expected outcomes, and novelty and feasibility scores. Novelty was assessed by a tool based on the results from the Semantic Scholar API. Idea 1 is described in Section S3, Idea 2 in Section S4, Idea 3 in Section S5, Idea 4 in Section S6, and Idea 5 in Section S7 (Supporting Information). The corresponding knowledge graphs are showing in Figure 8.

1	Research idea	Development of biomimetic microfluidic chips with enhanced heat transfer performance for biomedical applications
	Hypothesis	Integrating biomimetic materials, inspired by the lamellar structure of keratin scales, into microfluidic chips using soft lithography techniques will improve their mechanical behavior and heat transfer efficiency under cyclic loading conditions.
	Expected Outcomes	A 20-30% increase in heat transfer efficiency, a 15% reduction in failure rate under cyclic loading, and superior biocompatibility.
	Novelty/Feasibility	8/7
2	Research idea	Developing a novel collagen-based material with a hierarchical, interconnected 3D porous architecture to enhance crashworthiness, stiffness memory, and dynamic adaptability.
	Hypothesis	The hierarchical structure of collagen, when engineered into dynamic 3D architectures, can significantly improve these properties due to the interplay between biological interactions, cell signaling, and mechanical forces.
	Expected Outcomes	A 30% increase in crashworthiness, an 85% recovery rate of stiffness after deformation, a 25% increase in Young's modulus, and dynamic adaptability in response to biological and mechanical stimuli.
	Novelty/Feasibility	8/7
3	Research idea	Enhancing the mechanical properties of collagen-based scaffolds through a combination of tunable processability and nanocomposite integration adaptability.
	Hypothesis	optimizing material extrusion and electrospinning parameters, along with incorporating nanocomposites like graphene oxide, hydroxyapatite, and carbon nanotubes, will result in scaffolds with superior tensile strength, elasticity, and controlled pore sizes.
	Expected Outcomes	The expected outcomes include a 50% increase in tensile strength, a 40% improvement in elasticity, and enhanced base bite force metrics.
	Novelty/Feasibility	6/8
4	Research idea	Development of a novel biomimetic material by mimicking the hierarchical structure of nacre and incorporating amyloid fibrils.
	Hypothesis	The hierarchical structure of biomaterials, specifically nacre, enhances both superhydrophobic properties and mechanical robustness. By mimicking this structure and incorporating amyloid fibrils, advanced self-cleaning coatings with superior mechanical properties can be developed.
	Expected Outcomes	The expected outcomes include a water contact angle greater than 150 degrees, fracture toughness of at least 10 MPa $\sqrt{0.5}$, self-cleaning capabilities, and detailed AFM images showing the nanoscale hierarchical structure.
	Novelty/Feasibility	7/8
5	Research idea	Investigating the interaction between graphene and amyloid fibrils to create novel bioelectronic devices with enhanced electrical properties.
	Hypothesis	Binding of graphene to amyloid fibrils will result in a composite material with superior electrical conductivity and stability, which can be further optimized through engineered gene circuits that regulate the expression, secretion, and assembly of amyloid-forming proteins.
	Expected Outcomes	The expected outcomes include high-performance composite materials, detailed insights into binding mechanisms, optimized gene circuits, advanced bioelectronic devices, and broader scientific, technological, and societal impacts.
	Novelty/Feasibility	8/7

Planner**Overview of the Plan:**

1. Generate Random Keywords and Knowledge Path: Use the generate_path function to generate a knowledge path between two randomly selected keywords.
2. Define Terms and Relationships: The ontologist will define each term and discuss the relationships in the generated path.
3. Craft the Research Proposal: The scientist will craft a research proposal based on the definitions and relationships provided by the ontologist.
4. Expand the Research Proposal: Each specialized agent (hypothesisagent, outcomeagent, mechanismagent, designprinciplesagent, unexpectedpropertiesagent, comparisonagent, novelty_agent) will expand on their respective aspects of the research proposal.
5. Critique and Suggest Improvements: The critic_agent will summarize, critique, and suggest improvements to the research proposal.
6. Rate Novelty and Feasibility: Use the rate_novelty_feasibility function to rate the novelty and feasibility of the research idea.

Figure 9. The plan developed by the planner agent in response to the query from the user to generate research hypothesis from random keywords, as developed by the autonomous system. The process begins with the selection of random keywords, followed by the generation of a knowledge path that links the selected terms. Each term along the path is defined by an ontologist, who also elaborates on the relationships between them. Based on these definitions, a research proposal is crafted by a designated scientist. Subsequently, various specialized agents (hypothesis, outcome, mechanism, design principles, unexpected properties, comparison, and novelty agents) each expand upon their respective components of the proposal. The proposal is then critiqued by the critic_agent, who also suggests potential improvements. As the final step, the novelty and feasibility of the research proposal are assessed using a dedicated function, ensuring that the proposed ideas are both innovative and actionable.

(Supporting Information). The process demonstrates dynamic collaboration between the AI agents in constructing the research hypothesis. Initially, the planner proposes a comprehensive plan to accomplish the task, as shown in **Figure 9**. Following this, various agents execute the plan, starting with the generation of a knowledge graph, followed by the definition of key concepts and relationships by Ontologist agent. Scientist 1 then drafts the initial research proposal, which is further expanded by Scientist 2. Finally, the critic conducts a review, and the process concludes with an assessment of novelty and feasibility.

The randomly selected nodes for this experiment were “heat transfer performance” and “rhamphotheca” and the generated graph consists of concepts such as “lamellar structure,” “biomaterials,” “microfluidic chips,” “keratin scales,” and “biomimetic materials.” The proposed idea involves engineering the lamellar structure of biomaterials, inspired by keratin scales, into microfluidic chips using soft lithography techniques to improve their mechanical behavior and heat transfer efficiency under cyclic loading conditions. Expected outcomes of the resulting biomimetic microfluidic chips include a 20–30% increase in heat transfer efficiency compared to conventional microfluidic chips (the lamellar structure of the biomimetic materials will facilitate efficient heat dissipation), enhanced mechanical stability under cyclic loading conditions (the layered lamellar structure will provide enhanced mechanical strength and flexibility), with a failure rate reduced by 15%, and superior biocompatibility (due to the use of biocompatible materials), making them suitable for prolonged use in biomedical applications.

The design principles for biomimetic microfluidic chips focus on material selection, fabrication, integration, testing, biocompatibility, modeling, and optimization. Materials such as PDMS and hydrogels, which mimic the lamellar structure of keratin scales, are chosen for their biocompatibility and mechanical properties, with targeted thermal conductivity and Young’s modulus ranges. Soft lithography is employed for fabrication, opti-

mizing curing conditions and structural characterization. Integrating with microfluidic technology enhances heat transfer and mechanical stability, with design optimization via CAD and simulations. Testing includes mechanical and heat transfer assessments, while biocompatibility is evaluated through in vitro and in vivo tests. Finite Element Analysis (FEA) and Computational Fluid Dynamics (CFD) simulations help model heat transfer and fluid flow, guiding iterative design optimization based on performance metrics like heat transfer efficiency, mechanical stability, and biocompatibility.

Moreover, the model predicts that the biomimetic microfluidic chips may exhibit unexpected properties, such as self-healing capabilities, adaptive heat transfer, enhanced fluid dynamics, and improved chemical resistance. These properties are primarily attributed to the lamellar structure of the material, and the rationale behind them is summarized in **Table 5**.

For the proposed research idea, the critic agent summarizes the overall research hypothesis covering the key features and highlights strengths such as the innovative integration of biomimetic materials with microfluidic technology, detailed mechanisms for performance, and potential biomedical applications. It also acknowledges the exploration of self-healing and adaptive heat transfer. However, weaknesses include the complexity of the fabrication process, a lack of preliminary data, and concerns about long-term biocompatibility. To improve, the agent recommends conducting pilot studies, assessing scalability, and performing long-term biocompatibility testing. Moreover, the critic agent suggests the most impactful scientific questions with molecular modeling (How does the lamellar structure of biomimetic materials influence the heat transfer efficiency in microfluidic chips?) and synthetic biology and provides the pertinent key steps (Can biomimetic materials with a lamellar structure be engineered to exhibit self-healing properties under mechanical stress?). These specific directions can be used as

Table 5. Predicted unexpected properties for biomimetic microfluidic chips. The data summarizes the property, mechanism, and rationale.

Unexpected Property	Mechanism	Rationale
Self-Healing Properties	The lamellar structure might enable self-healing capabilities, where minor damages can be repaired autonomously, extending the lifespan of the chips.	The layered structure can facilitate the redistribution of stress and the healing of minor cracks, similar to natural biological systems.
Adaptive Heat Transfer	The heat transfer efficiency might adapt dynamically based on the thermal load, similar to natural biological systems.	The lamellar structure can facilitate dynamic adaptation to varying thermal loads, enhancing the overall thermal management capabilities.
Enhanced Fluid Dynamics	The lamellar structure might influence fluid dynamics within the microfluidic channels, leading to improved mixing and reduced pressure drop.	The layered structure can create micro-scale vortices and enhance fluid mixing, which is beneficial for applications requiring efficient mixing of reagents.
Improved Chemical Resistance	The lamellar structure might enhance the chemical resistance of the microfluidic chips, making them suitable for a wider range of applications.	The layered structure can act as a barrier to chemical penetration, protecting the underlying material from chemical degradation.

springboard for additional *in-situ* data collection; in the case of the modeling context, this can be implemented by incorporating a simulation engine, similar to what was done in recent work.^[38]

In the end, the assistant agent executes the tool to assess the novelty and feasibility of the proposed research idea against the literature. It then returns a detailed analysis as depicted in Figure 10 suggesting that the proposed research hypothesis has a high degree of novelty and a reasonable level of feasibility.

3. Conclusion

We introduced a multi-agent AI framework designed to autonomously generate and refine research hypotheses by leveraging LLMs and a comprehensive ontological knowledge graph 1, applied here in the context of biologically inspired materials. Our

results demonstrate the significant potential of integrating AI agents with specialized roles to tackle the complex and interdisciplinary nature of scientific discovery, particularly in the domain of bio-inspired materials. The automated system effectively navigated the intricate web of relationships within the knowledge graph, generating diverse and novel hypotheses that align with unmet research needs. The proposed approach, harnessing a modular, hierarchically organized (Figure 2) swarm of intelligence (Figure 1) similar to biological systems with multiple iterations to model the process of negotiating a solution during the process of thinking and reflecting about a problem, offers a much more nuanced reasoning approach than conventional zero-shot answers generated by AI systems, as shown in Figure 11.

The ontological knowledge graph representation of data plays a crucial role in our approach, as it serves as the foundational

Assistant

Novelty:

Score: 8/10

The concept of integrating biomimetic materials with microfluidic chips to enhance heat transfer and biocompatibility is relatively novel. The specific idea of using the lamellar structure inspired by keratin scales and engineering it into microfluidic chips using soft lithography techniques appears to be unique, as no direct matches were found in the literature. The existing literature does cover various aspects of microfluidic chip enhancements, including heat transfer, biocompatibility, and mechanical behavior, but the specific combination and approach proposed here seem to be unexplored.

Feasibility:

Score: 7/10

The feasibility of engineering lamellar structures inspired by keratin scales into microfluidic chips using soft lithography techniques is plausible. Soft lithography is a well-established method for fabricating microstructures, and biomimetic materials have been successfully integrated into various biomedical applications. However, the practical implementation of this specific structure and its performance under cyclic loading conditions would require thorough experimental validation. The complexity of achieving the desired mechanical behavior and heat transfer efficiency in a reliable and reproducible manner could pose challenges.

Figure 10. The results of the novelty and feasibility analysis as performed by the assistant agent for the “Biomimetic Microfluidic Chips” hypothesis, based on data collected using the Semantic Scholar API. As the analysis shows, the approach is considered unique due to its lack of direct matches in existing literature. Feasibility is evaluated based on the plausibility of implementing these structures using soft lithography, though challenges in mechanical behavior and heat transfer efficiency under cyclic loading were identified as potential hurdles requiring experimental validation.

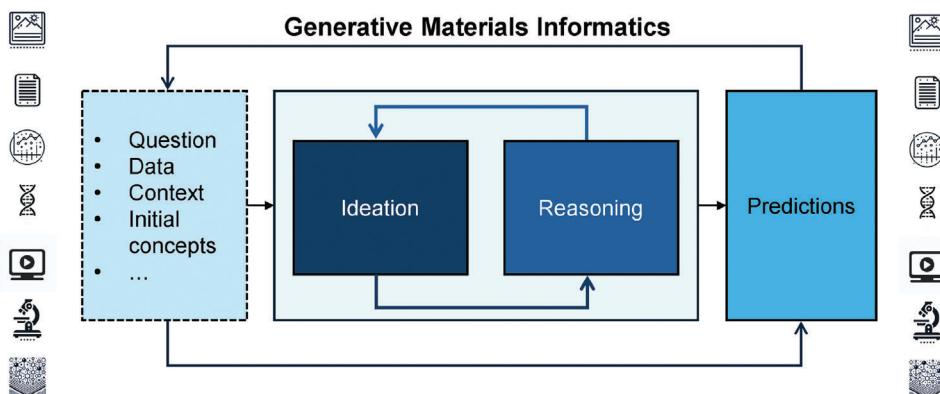


Figure 11. SciAgents presents a framework for generative materials informatics, showcasing the iterative process of ideation and reasoning driven by input data, questions, and context. The cycle of ideation and reasoning leads to predictive outcomes, offering insights into new material designs and properties. The visual elements on the edges represent various data modalities such as images, documents, scientific data, DNA sequences, video content, and microscopy, illustrating the diverse sources of information feeding into this process.

structure that guides the research idea generation, ensuring that the hypotheses proposed by the AI agents are both informed by and rooted in a vast network of interconnected scientific concepts. By systematically navigating this graph, our multi-agent system identifies and capitalizes on previously unrecognized relationships, aiming towards the creation of highly-rated innovative ideas that are as feasible as they are groundbreaking. The incorporation of assessment strategies is an important strategic aspect that reflects adversarial relationships commonly identified in conventional research strategies, such as team-based efforts or peer-review. A notable feature was the finding that the autonomous multi-agent system can develop sophisticated problem solving strategies (see, Figure 7) on its own. These types of results are expected to improve as more powerful foundation models become available, especially with better long-term planning and reasoning capabilities.

The multi-agent approach proved particularly effective in decomposing the scientific discovery process into manageable sub-tasks, enabling a more systematic exploration of the knowledge landscape. By assigning distinct roles to each agent—ranging from path generation and deep analysis to hypothesis formulation and critical review, we achieved a thorough and rigorous development of research ideas. Our experiments showed that the system could consistently produce hypotheses with high novelty and feasibility, supported by contextually enriched data and iterative feedback mechanisms that mirrored traditional scientific methodologies. The incorporation of specific priority modeling and simulation tasks, for instance, offers direct pathways to incorporate additional mechanisms to solicit new physics-based data (e.g., by running Density Functional Theory models, molecular dynamics, finite element/difference solvers, etc.).^[38,39] As such, the approach presented here offers significant potential in not only developing research questions but also expanding the set of first-principles sourced data. If deployed at scale, this can aid our quest to generate large materials-focused datasets strategically expanding beyond what is currently known. Based on the execution efficacy, it is possible to generate thousands or tens of thousands of individual results within days, which if filtered by a set of criteria (e.g., novelty, feasibility, or how well it meets a target) can gener-

ate a high-efficacy innovation framework for generative materials informatics.

One of the key contributions of this study is the demonstration of how AI-driven agents can autonomously generate, critique, and refine scientific hypotheses, offering a scalable and efficient alternative to conventional research approaches. The integration of tools to assess novelty against existing literature further strengthens the validity of the generated hypotheses, ensuring that the system not only produces innovative ideas but also eliminates redundancies with prior research. This capability positions the system as a powerful tool for accelerating discovery and fostering cross-disciplinary innovation.

In fields such as biological materials analysis, identifying common mechanisms that hold for a variety of systems and that can be applied to solve challenging engineering problems, remains a major challenge. This work underscores the potential of generative AI in potentially scaling the scientific process, opening new avenues for exploration and discovery across various fields of study. As we can automate, and hence accelerate the generation of research ideas, this multi-agent system paves the way for a future where AI could then contribute as an integral player in shaping the direction and pace of scientific advancement.

While SciAgents demonstrates substantial promise in generating novel research ideas, questions remain regarding the feasibility of certain proposals. In this study, we have taken initial steps to address these limitations by incorporating additional layers of support. Specifically, AI agents refined proposals through iterative feedback loops, enriching the context to better align with established methodologies. Additionally, we evaluated the novelty of research ideas by using Semantic Scholar API calls to compare generated concepts against existing knowledge, providing an initial metric for assessing uniqueness and relevance. However, ensuring the robustness of these ideas will require further exploration. The flexibility of the multi-agent system supports continuous refinement and adaptation, which can improve both feasibility and applicability over time. For instance, integrating dynamic collaboration between AI and human experts—particularly

experimentalists—could significantly strengthen the feasibility of research ideas. Though not fully explored in this work, this AI-human collaboration remains an exciting avenue for future research. Additionally, ongoing advancements in large language models are anticipated to enhance both the novelty and feasibility of generated ideas, and these improvements can be seamlessly integrated into the SciAgents framework to keep pace with technological progress. For instance, we have included an example of a research idea developed by SciAgents using the o1-preview model in Section S8 (Supporting Information), demonstrating higher reasoning capabilities and more impactful results compared to GPT-4.

SciAgents, in its current form, excels at generating innovative research ideas in biologically inspired materials, offering valuable starting points for materials scientists, experimentalists, and biologists to explore further. These ideas provide promising directions for interdisciplinary research, serving as a foundation for new scientific insights and collaborations. To further expand the system's impact, future work could focus on implementing mechanisms for validating research hypotheses. One direction could involve adding agents capable of performing computational simulations or extracting insights from experimental data, directly addressing the need for robust hypothesis validation. Furthermore, establishing clear metrics—such as efficiency and accuracy—alongside a collaborative evaluation framework involving human experts would allow for a more comprehensive assessment of SciAgents' effectiveness and applicability (similar to Ref. [27], where human experts were involved in evaluating NLP-related research ideas). The system's modular, human-in-the-loop design provides the flexibility to incorporate these enhancements, allowing for the seamless integration of validation processes and expert feedback. Through these advancements, SciAgents could evolve into an increasingly robust tool that not only generates but rigorously evaluates ideas, driving research toward impactful real-world applications. We leave these exciting developments as directions for future work.

With the integration of pertinent contextual data and domain-specific knowledge graphs, SciAgents has the potential to be adapted for a range of scientific domains and subfields, including chemistry, metallic alloys, semiconductors, and physics. A possible future development could involve a universal knowledge graph, connecting concepts across all scientific domains. Such an integration would allow SciAgents to bridge ideas between fields, fostering interdisciplinary research and accelerating innovation by uncovering insights that span multiple scientific disciplines. This could be further strengthened by integrating fine-tuned LLMs for different fields or newly developed models (e.g., o1-preview), enabling more specialized reasoning across diverse areas and enhancing the quality and relevance of the generated insights. Additionally, collaboration with human experts would provide critical oversight, ensuring that AI-generated ideas are grounded in real-world applicability and guiding the research toward practical solutions. Hence, we believe that the framework presented here offers a blueprint for next-generation of AI-driven research tools, capable of synthesizing vast amounts of data into actionable insights, ultimately leading to breakthroughs that might otherwise remain undiscovered.

4. Experimental Section

Ontological Knowledge Graph: A large graph generated as part of earlier work^[6] was used in this research. The graph utilized here includes 33 159 nodes and 48 753 edges and represents the giant component of the graph generated from around 1000 papers with 92 communities. The BAAI/bge-large-en-v1.5 embedding model was used.

Heuristic Pathfinding Algorithm with Random Waypoints: The algorithm presented in this work combines heuristic-based pathfinding with node embeddings and randomized waypoints to discover diverse paths in a graph. The primary goal is to find a path between a source and a target node by estimating distances using node embeddings. The embeddings are generated using a pre-trained model and are crucial for the heuristic function, which estimates the distance between the current node and the target node. By relying on these embeddings, the algorithm adapts to the topological structure of the graph, allowing it to effectively traverse complex networks. Additionally, the algorithm uses a modified version of Dijkstra's algorithm that introduces a randomness factor to the priority queue, creating paths that are not strictly deterministic.^[46] The randomness factor was chosen to be 0.2 in the experiments.

An additional feature of the algorithm is the introduction of random waypoints to diversify the pathfinding process. These waypoints are selected from neighboring nodes that are not part of the initial path, enabling the algorithm to explore alternative routes. The randomization factor controls the balance between heuristic-driven search and stochastic exploration, making it flexible for different use cases. After the path is found, a subgraph consisting of the path nodes and their second-hop neighbors is generated, providing a broader context for the discovered route. The resulting paths are then used as substrate for graph reasoning.

The overall approach is as follows:

Heuristic Pathfinding with Randomization and Waypoints

Input:

- Graph G
- Embedding tokenizer and model
- Source and target nodes
- Node embeddings E
- Randomness factor α
- Number of random waypoints k

Computation of output: Path P from source to target, subgraph G' , shortest path length

1. **Initialize:** Set $P = []$, priority queue $Q = [(0, \text{source})]$, visited nodes $V = \{\}$
2. **Find closest nodes:** Use embedding tokenizer and model to find best-fitting nodes for source and target.
3. **Estimate heuristic:** Compute distance between current node and target using embeddings.
4. **Randomized Dijkstra:**
 - (a) Use Dijkstra's algorithm, adding a random factor α to prioritize exploration over purely heuristic pathfinding.
 - (b) **While** Q is not empty:
 - Pop node u with the lowest cost from Q
 - **If** $u = \text{target}$: Return path P
 - Mark u as visited
 - **For each neighbor** v of u :
 - Calculate heuristic distance $h(v, \text{target})$ using embeddings
 - Compute cost of visiting v as $\text{cost}(v) = h(v, \text{target}) + \alpha \times \text{random}()$
 - **If** v not in V : Add $(\text{cost}(v), v)$ to Q
 5. **Add random waypoints:**
 - Randomly select waypoints from neighbors of nodes in P , ensuring they are not already in the path.

- For each waypoint, compute shortest path to the next waypoint and extend P .
- 6. **Return path:** After waypoints, compute the final leg from the last waypoint to target.
- 7. **Build subgraph:** Create a subgraph G' containing all nodes and edges along the path.
- 8. **Save results:** If enabled, save the path visualization and subgraph to HTML and GraphML files.
- 9. **Return:** Path P , subgraph G' , shortest path length.

Graph Reasoning—Initial Ideation: The initial step in the approach develops a scientific hypothesis based on a knowledge graph derived from a heuristic path in a given graph G as described in Section 4.2. Here the graph G represents a set of interconnected nodes, where each node can represent an entity or concept, and edges represent relationships between these nodes. The algorithm begins by identifying two key nodes, keyword_1 and keyword_2, which can either be explicitly specified or randomly selected from G . If the shortest_path flag is set to True, the function computes the shortest path between these nodes by using embeddings to estimate the best-fitting nodes, leveraging a pre-existing function called find_path. If shortest_path is set to False, a heuristic pathfinding approach is employed, which incorporates randomization and potentially random waypoints to explore more diverse paths. The graph structure is used not only for identifying the connectivity between the nodes but also for guiding the algorithm's search for the most relevant or exploratory paths based on the node embeddings.

Once a path between keyword_1 and keyword_2 is established, the function constructs a knowledge graph from the path and its relationships. This knowledge graph consists of the nodes traversed and the relationships (edges) between them. The graph's structure is vital as it is used to form the input for a generative model, which expands on the graph's nodes and relationships by providing definitions and explanations. The function also generates a novel research hypothesis by analyzing the graph, synthesizing a hypothesis based on the relationships and concepts discovered along the path. The structure of the graph helps to frame the scientific inquiry, with the hypothesis leveraging the graph's connections to predict outcomes, explore mechanisms, and propose innovative ideas. This output is formatted as a JSON object with fields like "hypothesis," "outcome," and "design_principles," each reflecting different aspects of the potential research grounded in the graph's topology.

A key aspect of the process is the use of natural language generation to dynamically expand on the concepts represented by the nodes and edges of the knowledge graph. For each node, the generative model provides detailed definitions and explanations of the scientific concepts it represents. The relationships between the nodes, represented by the edges, are also expanded to give context to how these concepts are interconnected. This approach not only builds a deeper understanding of individual components of the graph but also enhances the user's ability to interpret the complex interrelations between them, thereby setting the foundation for novel scientific inquiry. The response generated by the model includes comprehensive descriptions of these relationships, ensuring that the resulting graph becomes a robust substrate for knowledge synthesis.

After the knowledge graph is expanded, the algorithm generates a structured scientific hypothesis that leverages each of the nodes and relationships in the graph. The output, in JSON format, provides key fields such as "mechanisms," "unexpected_properties," and "comparison," offering a highly detailed analysis. The "mechanisms" field discusses predicted chemical, biological, or physical interactions, while "unexpected_properties" anticipates emergent behaviors from novel combinations of concepts in the graph. This comprehensive hypothesis formulation process allows for the exploration of unexplored areas of study, providing an innovative and grounded approach to scientific discovery based on the structure of the graph and its conceptual relationships.

Expansion of the Initial Concepts: The final phase of the methodology focuses on leveraging the expanded research concept to identify key scientific questions and prioritize actionable research directions, particularly in the domains of molecular modeling and synthetic biology. This phase employs a generative model to analyze the complete research document, which includes the knowledge graph, expanded concepts, and critical reviews, with the goal of extracting the most impactful scientific questions. These questions are then further expanded into detailed experimental and simulation plans, or other specific aspects that a user wants to explore in detail.

Using the JSON developed as described in Section 4.3.1 several systematic steps are conducted.

Step 1: Prompt-Driven Expansion of Key Research Aspects

The next phase involves systematically expanding specific aspects of the hypothesis using a series of targeted prompts. For each aspect of the research, a detailed prompt is constructed to critically assess and improve the scientific content of that aspect. The primary aspects are drawn from the JSON dictionary, where all elements were iterated in that data structure.

The following steps summarize how the model expands each research aspect:

- A prompt is created for each field in the JSON data structure, asking the model to **expand upon the original content** by adding quantitative details such as chemical formulas, material properties, or specific experimental methods.
The model is also instructed to provide a step-by-step rationale for the proposed scientific improvements.
- For example, the prompt format includes:
Expand on the following aspect: field.
Critically assess the original content, add specifics, such as chemical formulas, sequences, microstructures, and rational improvements:
JSON_dictionary[field]
- The model generates expanded content under a heading such as ### Expanded Mechanisms or ### Expanded Outcomes. Each response is added to res_data_expanded to track the expanded fields.
- The iterative process is repeated for each of the first seven fields in res_data, ensuring that every major aspect of the research concept is thoroughly evaluated and improved upon.

Step 2: Compilation and Summary of Expanded Content

After the expansion phase, the system compiles the results into a structured document, starting with the original knowledge graph and hypothesis, followed by the expanded research aspects. This forms a cohesive research narrative. The complete document includes sections such as:

```
# Research concept between {start_node} and {end_node}
### KNOWLEDGE GRAPH:
{path_string}

### EXPANDED GRAPH:
{res_data['expanded']}

### PROPOSED RESEARCH:
{formatted_text}

### EXPANDED DESCRIPTIONS:
{expanded_text}
```

Step 3: Scientific Critique and Review

Following the expansion, a prompt is issued to the model to **critically review the entire document**. The review is designed to evaluate both the strengths and weaknesses of the proposed research and to suggest improvements. This step is crucial in ensuring that the expanded content is scientifically rigorous and logical. The prompt asks for:

Provide a thorough critical scientific review with strengths, weaknesses, and suggested improvements.

You are a critical AI assistant collaborating with a group of scientists to assess the potential impact of a research proposal. Your primary task is to evaluate a proposed research hypothesis for its novelty and feasibility, ensuring it does not overlap significantly with existing literature or delve into areas that are already well-explored. You will have access to the Semantic Scholar API, which you can use to survey relevant literature and retrieve the top 10 results for any search query, along with their abstracts. Based on this information, you will critically assess the idea, rating its novelty and feasibility on a scale from 1 to 10 (with 1 being the lowest and 10 the highest). Your goal is to be a stringent evaluator, especially regarding novelty. Only ideas with a sufficient contribution that could justify a new conference or peer-reviewed research paper should pass your scrutiny.

After careful analysis, return your estimations for the novelty and feasibility rates. If the tool call was not successful, please re-call the tool until you get a valid response.

After the evaluation, conclude with a recommendation and end the conversation by stating "TERMINATE".

Figure 12. The profile of the novelty assistant LLM agent implemented in the automated multi-agent approach for rating the novelty of the research idea.

The result is a critical review that is appended to the final document as ‘‘SUMMARY, CRITICAL REVIEW, AND IMPROVEMENTS’’.

Step 4: Identification of Modeling and Experimental Priorities

Finally, the model is prompted to identify the **most impactful scientific questions** related to molecular modeling and synthetic biology.

Separate prompts are issued for each domain, asking the model to:

- Identify a key research question that can be tackled using **molecular modeling**, and outline steps to conduct such modeling, including any specific tools or techniques.
- Similarly, for **synthetic biology**, the model is prompted to outline an experimental plan, detailing unique aspects such as gene-editing protocols, biological sequences, or organism-specific techniques.

Examples of these prompts:

Identify the single most impactful scientific question that can be tackled with molecular modeling. Outline key steps for conducting such modeling.

Identify the most impactful question for synthetic biology and provide an experimental setup.

The responses are appended to the final document under ‘‘MODELING AND SIMULATION PRIORITIES’’ and ‘‘SYNTHETIC BIOLOGY EXPERIMENTAL PRIORITIES’’.

Final Document and Output

The entire research concept, expanded and reviewed, is then compiled into a final document which is saved as both a PDF and CSV file for further analysis. The final document contains:

- The original knowledge graph and proposed research hypothesis.
- Expanded descriptions of key research aspects.
- A critical review of the proposal.
- Research priorities for molecular modeling and synthetic biology.

This provides a comprehensive output that transitions the generated hypothesis into a detailed, actionable research plan.

Agentic Modeling: AI agents were designed using the general-purpose LLM GPT-4 family models. The automated multi-agent collaboration is implemented in the AutoGen framework,^[47] an open-source ecosystem for agent-based AI modeling.

In our multi-agent system, the human agent is constructed using UserProxyAgent class from Autogen, and Assistant, Planner, Ontologist, Scientist 1, Scientist 2, and Critic agents are created via AssistantAgent class from Autogen; and the group chat manager is created using GroupChatManager class. Each agent is assigned a role through a profile description included as system_message at their creation. The full profile of the agents is provided in Figure S2 for the planner, Figure S3 for the assistant,

Figure S4 for the Ontologist, Figure S5 for the Scientist 1, Figure S6 for the Scientist 2, and Figure S7 (Supporting Information) for the Critic.

Function and Tool Design: All the tools implemented in this work are defined as python functions. Each function is characterized by a name, a description, and input properties which have a proper description.

Semantic Scholar Analysis: The Semantic Scholar API was used, an AI-powered search engine for academic resources, to search for related publications using a set of keywords. To ensure a thorough assessment of the research idea, a tool featuring an AI agent named the “novelty assistant” was implemented, which calls the Semantic Scholar API three times using different combinations of keywords selected based on the research hypothesis. The profile of this agent is shown in **Figure 12**. For each function call, the ten most relevant publications are returned, including their titles and abstracts. The novelty assistant agent then thoroughly analyzes the abstracts and provides a review describing the novelty of the research idea.

Audio Representation: Several examples of files created using the PDF2Audio tool (<https://github.com/lamm-mit/PDF2Audio>). The audio simulates a podcast-style discussion between two people about SciAgents, based on the content of the manuscript and a material design developed by the algorithm.

This tool facilitates the transformation of PDF files into audio outputs, designed to cater to diverse presentation formats such as podcasts, lectures, and summaries. The process begins with PDF file upload, from which text is extracted using a PDF reader. Users can customize text generation through instruction templates tailored for specific use cases, such as podcasts or summaries, which guide content style, structure, and tone. Text is generated using OpenAI’s o1-preview model, selected for their capacity to handle large language tasks with nuanced control over tone and style.

The generated text undergoes a natural language synthesis phase using text-to-speech (TTS) models, here tts-1, which deliver high-definition audio output in voices specified by the user. Voice selection is further enhanced through predefined voice options to match different stylistic needs. The algorithm allows for iterative text refinement through a feedback loop, integrating user comments and transcript edits in subsequent runs. The final audio file is delivered as an MP3 format, with transcripts and the original text provided for reference.

In addition to podcasts and lecture-style formats, the tool supports a variety of audio content styles including scientific summaries, conversational panels, and multilingual outputs. Each format employs customized instruction templates, defining tone, structure, and detail level based on the desired style. For example, the SciAgents material discovery summary format creates an engaging dialogue between an expert and a student to explain scientific concepts, while the lecture format generates structured monologues for educational content. Templates are further customizable for language-specific instructions, enabling formats in French, German, Spanish, and other languages, with specialized directives to ensure accessible explanations and cultural relevance.

This flexibility allows users to tailor outputs for diverse audiences and applications, from academic discussions to general public outreach.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors acknowledge support from USDA (2021-69012-35978), DOE-SERDP (WP22-S1-3475), ARO (79058LSCSB, W911NF-22-2-0213 and W911NF2120130) as well as the MIT-IBM Watson AI Lab, MIT's Generative AI Initiative, and Google. Additional support from NIH (U01EB014976 and R01AR077793) is acknowledged. A.G. gratefully acknowledges the financial support from the Swiss National Science Foundation (project #P500PT_214448).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article. Codes used for this study can be found on GitHub at <https://github.com/lamm-mit/SciAgentsDiscovery> and <https://github.com/lamm-mit/GraphReasoning/>.

Keywords

bio-inspired materials, biological design, knowledge graph, large language model, materials design, multi-agent system, natural language processing, scientific AI

Received: September 9, 2024

Revised: November 12, 2024

Published online:

- [1] T. van der Zant, M. Kouw, L. Schomaker, *Studies in Applied Philosophy, Epistemology and Rational Ethics* **2013**, 5, 107.
- [2] K. Guo, Z. Yang, C.-H. Yu, M. J. Buehler, *Mater. Horiz.* **2021**, 8, 1153.
- [3] Y. Liu, T. Zhao, W. Ju, S. Shi, *J. Materiomics* **2017**, 3, 159.
- [4] Y. Hu, M. J. Buehler, *APL Mach. Learn.* **2023**, 1, 010901.
- [5] M. Matsumoto, *MRS Bull.* **2022**, 1.
- [6] M. J. Buehler, *Mach. Learn.: Sci. Technol.* **2024**.
- [7] Q. Zhang, Y. Hu, J. Yan, H. Zhang, X. Xie, J. Zhu, H. Li, X. Niu, L. Li, Y. Sun, W. Hu, *Adv. Mater.* **2024**, 36, 2405163.
- [8] C. Lu, C. Lu, R. T. Lange, J. Foerster, J. Clune, D. Ha, *arXiv preprint arXiv:2408.06292* **2024**.
- [9] W. L. Ng, G. L. Goh, G. D. Goh, J. S. J. Ten, W. Y. Yeong, *Adv. Mater.* **2024**, 36, 2310006.
- [10] G. Lei, R. Docherty, S. J. Cooper, *Digital Discovery* **2024**.
- [11] OpenAI, *arXiv:2303.08774* **2024**.
- [12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, *Adv. Neural Inform. Process. Syst.* **2017**, 30.
- [13] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean, Q. Fedus, *arXiv preprint arXiv:2206.07682* **2022**.
- [14] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, et al., *arXiv preprint arXiv:2307.09288* **2023**.
- [15] T. Teubner, C. M. Flath, C. Weinhardt, W. van der Aalst, O. Hinz, *Busin. Inform. Syst. Eng.* **2023**, 65, 95.
- [16] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, Y. Du, C. Yang, Y. Chen, Z. Chen, J. Jiang, R. Ren, Y. Li, X. Tang, Z. Liu, P. Liu, J. Nie, J. Wen, *arXiv preprint arXiv:2303.18223* **2023**.
- [17] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, P. Barham, H. W. Chung, C. Sutton, S. Gehrmann, P. Schuh, K. Shi, S. Tsvyashchenko, J. Maynez, A. Rao, P. Barnes, Y. Tay, N. Shazeer, V. Prabhakaran, E. Reif, N. Du, B. Hutchinson, R. Pope, J. Bradbury, J. Austin, M. Isard, G. Gur-Ari, P. Yin, T. Duke, A. Levskaya, et al., *J. Mach. Learn. Res.* **2023**, 24, 1.
- [18] S. Gunasekar, Y. Zhang, J. Aneja, C. C. T. Mendes, A. Del Giorno, S. Gopi, M. Javaheripi, P. Kauffmann, G. de Rosa, O. Saarikivi, A. Salim, S. Shah, H. S. Behl, X. Wang, S. Bubeck, R. Eldan, A. T. Kalai, Y. T. Lee, Y. Li, *arXiv preprint arXiv:2306.11644* **2023**.
- [19] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. d. I. Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M. Lachaux, P. Stock, T. Le Scao, T. Lavril, T. Wang, T. Lacroix, W. El Sayed, *arXiv preprint arXiv:2310.06825* **2023**.
- [20] K. Girotra, L. Meincke, C. Terwiesch, K. T. Ulrich, *Available at SSRN* 4526071 **2023**.
- [21] M. J. Buehler, *ACS Engineering AU* **2023**.
- [22] K. M. Jablonka, Q. Ai, A. Al-Feghali, S. Badhwar, J. D. Bocarsly, A. M. Bran, S. Bringuier, L. C. Brinson, K. Choudhary, D. Cirić, S. Cox, W. A. de Jong, M. L. Evans, N. Gastello, J. Genzling, M. V. Gil, A. K. Gupta, Z. Hong, A. Imran, S. Kruschwitz, A. Labarre, J. Lala, T. Liu, S. Ma, S. Majumdar, G. W. Merz, N. Moitessier, E. Moubarak, B. Mourino, B. Pelkie, et al., *Digit. Discov.* **2023**, 2, 1233.
- [23] A. M. Bran, S. Cox, O. Schilter, C. Baldassari, A. D. White, P. Schwaller, *Nat. Mach. Intell.* **2024**, 1.
- [24] R. K. Luu, M. J. Buehler, *Adv. Sci.* **2024**, 11, 2306724.
- [25] W. Lu, R. K. Luu, M. J. Buehler, *arXiv:2409.03444* **2024**.
- [26] M. J. Buehler, *Adv. Funct. Mater.* n/a, 2409531.
- [27] C. Si, D. Yang, T. Hashimoto, *arXiv preprint arXiv:2409.04109* **2024**.
- [28] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, *Adv. Neural Inf. Process. Syst.* **2022**, 35, 24824.
- [29] J. White, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, D. C. Schmidt, *arXiv preprint arXiv:2302.11382* **2023**.
- [30] Y. Zhou, A. I. Muresanu, Z. Han, K. Paster, S. Pitis, H. Chan, J. Ba, *arXiv preprint arXiv:2211.01910* **2022**.
- [31] J. Sun, C. Xu, L. Tang, S. Wang, C. Lin, Y. Gong, H.-Y. Shum, J. Guo, *arXiv preprint arXiv:2307.07697* **2023**.
- [32] P. Shetty, A. C. Rajan, C. Kuenneth, S. Gupta, L. P. Panchumartti, L. Holm, C. Zhang, R. Ramprasad, *npj Comput. Mater.* **2023**, 9, 52.
- [33] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, X. Wu, *IEEE Transactions on Knowledge and Data Engineering* **2024**.
- [34] J. Dagdelen, A. Dunn, S. Lee, N. Walker, A. S. Rosen, G. Ceder, K. A. Persson, A. Jain, *Nat. Commun.* **2024**, 15, 1418.
- [35] M. Schilling-Wilhelmi, M. Ríos-García, S. Shabih, M. V. Gil, S. Miret, C. T. Koch, J. A. Márquez, K. M. Jablonka, *arXiv preprint arXiv:2407.16867* **2024**.
- [36] B. Ni, M. J. Buehler, *Extreme Mechanics Letters* **2024**, 67, 102131.
- [37] I. Stewart, M. Buehler, *ChemRxiv* **2024**.
- [38] A. Ghafarollahi, M. J. Buehler, *Digital Discovery* **2024**, 1389.
- [39] A. Ghafarollahi, M. J. Buehler, *arXiv preprint arXiv:2407.10022* **2024**.

- [40] T. Giesa, D. Spivak, M. Buehler, *Adv. Eng. Mater.* **2012**, *14*, 9.
- [41] D. Spivak, T. Giesa, E. Wood, M. Buehler, *PLoS ONE* **2011**, *6*, 9.
- [42] Q. Wu, G. Bansal, J. Zhang, Y. Wu, S. Zhang, E. Zhu, B. Li, L. Jiang, X. Zhang, C. Wang, *arXiv preprint arXiv:2308.08155* **2023**.
- [43] Y. Wu, T. Yue, S. Zhang, C. Wang, Q. Wu, *arXiv:2403.11322*, **2024**.
- [44] Y. Wu, F. Jia, S. Zhang, H. Li, E. Zhu, Y. Wang, Y. T. Lee, R. Peng, Q. Wu, C. Wang, in *ArXiv preprint arXiv:2306.01337*, **2023**.
- [45] OpenAI API, <https://openai.com/blog/openai-api>.
- [46] E. W. Dijkstra, *Numerische Mathematik* **1959**, *1*, 269.
- [47] Q. Wu, G. Bansal, J. Zhang, Y. Wu, B. Li, E. Zhu, L. Jiang, X. Zhang, S. Zhang, J. Liu, A. Awadallah, R. W. White, D. Burger, C. Wang, *arXiv:2308.08155* **2023**.