

Generative Retrieval-Augmented Ontologic Graph and Multiagent Strategies for Interpretive Large Language Model-Based Materials Design

Published as part of ACS Engineering Au virtual special issue "Materials Design".

Markus J. Buehler*



Cite This: ACS Eng. Au 2024, 4, 241–277



Read Online

ACCESS |



Metrics & More



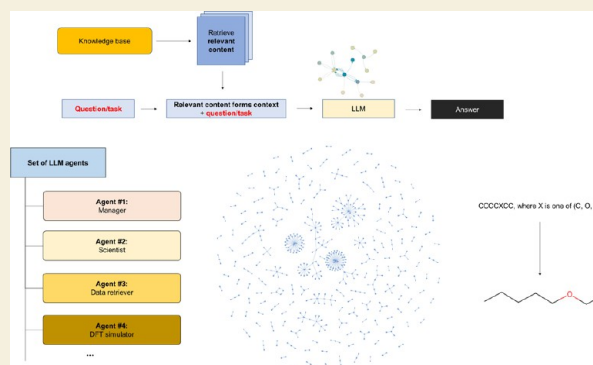
Article Recommendations



Supporting Information

ABSTRACT: Transformer neural networks show promising capabilities, in particular for uses in materials analysis, design, and manufacturing, including their capacity to work effectively with human language, symbols, code, and numerical data. Here, we explore the use of large language models (LLMs) as a tool that can support engineering analysis of materials, applied to retrieving key information about subject areas, developing research hypotheses, discovery of mechanistic relationships across disparate areas of knowledge, and writing and executing simulation codes for active knowledge generation based on physical ground truths. Moreover, when used as sets of AI agents with specific features, capabilities, and instructions, LLMs can provide powerful problem-solution strategies for applications in analysis and design problems. Our experiments focus on using a fine-tuned model, MechGPT, developed based on training data in the mechanics of materials domain. We first affirm how fine-tuning endows LLMs with a reasonable understanding of subject area knowledge. However, when queried outside the context of learned matter, LLMs can have difficulty recalling correct information and may hallucinate. We show how this can be addressed using retrieval-augmented Ontological Knowledge Graph strategies. The graph-based strategy helps us not only to discern how the model understands what concepts are important but also how they are related, which significantly improves generative performance and also naturally allows for injection of new and augmented data sources into generative AI algorithms. We find that the additional feature of relatedness provides advantages over regular retrieval augmentation approaches and not only improves LLM performance but also provides mechanistic insights for exploration of a material design process. Illustrated for a use case of relating distinct areas of knowledge, here, music and proteins, such strategies can also provide an interpretable graph structure with rich information at the node, edge, and subgraph level that provides specific insights into mechanisms and relationships. We discuss other approaches to improve generative qualities, including nonlinear sampling strategies and agent-based modeling that offer enhancements over single-shot generations, whereby LLMs are used to both generate content and assess content against an objective target. Examples provided include complex question answering, code generation, and execution in the context of automated force-field development from actively learned density functional theory (DFT) modeling and data analysis.

KEYWORDS: hierarchical, biomaterials, generative autoregressive transformer, multiscale modeling, materials by design, language models



1. INTRODUCTION

Large language models (LLMs) based on the general framework of decoder-only autoregressive transformer models^{1–6} provide powerful tools for scientific exploration. The specific class of Generative Pretrained Transformer (GPT) models has received significant attention across many fields of inquiry, suggesting remarkable possibilities, especially for generative forward and inverse tasks.^{5,7–15} However, our understanding of their behavior is still in its infancy, and issues remain with respect to fact recall and potential hallucination, which requires careful validation of their predictions and a thorough exploration of implications.^{16–23} There have been several recent developments

that propose the use of LLM-type models,^{24–26} and more generally attention-based transformer architectures, to capture the behavior of physical systems, including materials.^{27–30} Other studies have suggested the broad applicability of transformer

Received: September 24, 2023

Revised: December 6, 2023

Accepted: December 7, 2023

Published: January 12, 2024



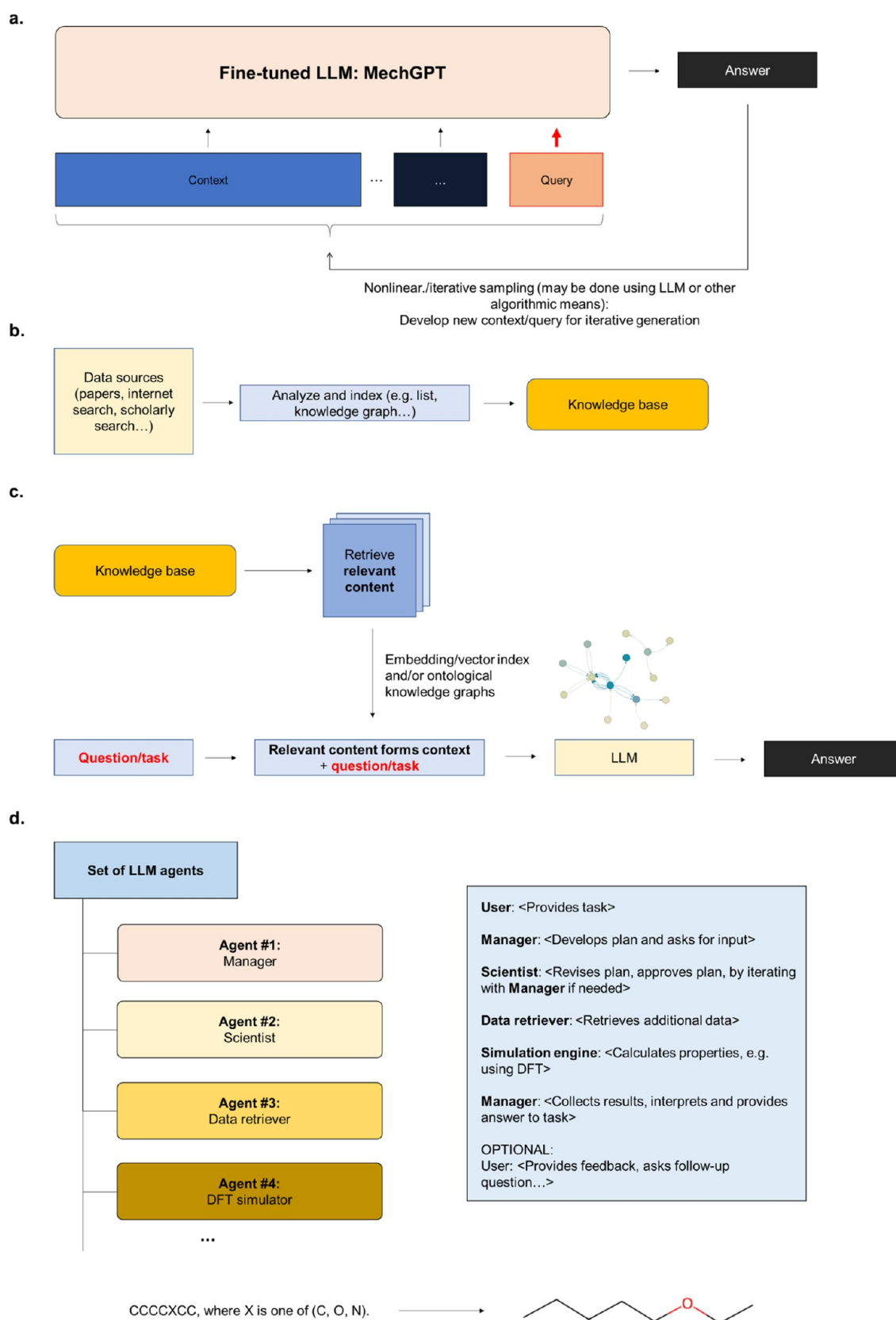


Figure 1. Summary of the strategies used in this study, using Large Language Models (LLMs) to solve various tasks ranging from question answering and Ontological Knowledge Graph construction to multi-AI agent conversation for group problem solving. Panel a visualizes a general perspective of how LLMs use context and queries to provide an answer. In a conventional setting, LLMs are queried against their parameter-based knowledge. By developing a knowledge base from data sources (panel b), we can augment the response of an LLM by providing relevant context retrieved from the knowledge base with the question to provide an answer, as visually summarized in panel c. Other mechanisms for sampling discussed in this paper include nonlinear sampling strategies. Knowledge sources can be developed using a variety of strategies, for instance, based on a literature search (e.g.,

Figure 1. continued

identifying search terms and downloading highly cited or recent papers). (d) Agent-based LLM model where multiple LLMs, each with special skills and knowledge or retrieval powers, interact. In the example sketched out here, we consider a set of 4 agents: a manager that oversees the task solution process, a scientist with a broad background and review specialization, a data retriever that can retrieve data (e.g., from the Internet, scientific papers, or code), and a simulation engine (e.g., a DFT simulator). The right panel shows a conceptual conversation carried out by these agents to solve a problem. As illustrated in the paper, LLMs can also autonomously write and execute code to solve complex materials modeling and design problems, such as force-field development.

models in use cases ranging from protein folding,³¹ protein property prediction,^{32–36} mechanical field predictions,^{37,38} as optimizers,³⁹ materials design,^{14,28,29,40,41} and as educational tools,^{42–44} among many others. The availability of large models is facilitated by recent developments such as the GPT-4 model,⁴⁵ and also releases of open-source LLMs such as Llama-2⁴⁶ or Falcon.^{47,48} Here, we discuss strategies for how LLMs can be improved to provide more accurate responses, especially in the context of materials analysis and design, and what mechanisms we can use to elicit more nuanced, detailed, and relevant outcomes, including interpretability. The strengths of such models reside in part due to their intrinsic ability to learn structurally, as transformer models are ultimately graph-forming models, and when combined into sets of autonomously interacting agents, they can form deep networks of interacting systems to solve complex problems that can not only retrieve baked-in information from training, but retrieve knowledge from sources, Internet searches, and also write code to run first-principles atomistic simulations by connecting disparate simulation engines to collect new data.⁴⁹ The intrinsic graph-forming capacity can thereby be augmented by amending such mechanistic strategies at higher levels in the materiomics space of structure–property relationships by discovering and using ontological principles akin to categorization.^{50–58} This then provides a fertile environment to solve a variety of materials analysis problems, including design.

We build this work on a recently developed MechGPT model^{30,59} that was trained on a data set of domain knowledge in the area of mechanics of materials but was extended here by including a larger training set to also feature Wikipedia articles related to mechanics content as well as textbooks on materials failure⁶⁰ (for details, see [Materials and Methods](#)). The purpose of this study is to explore the behavior of the model in its trained form and enhance generative capabilities by using complex prompting to elicit more accurate responses, more context, and featuring higher levels of details. We also discuss the role of nonlinear prompting strategies where LLMs are used to generate concepts, then judge these concepts against a task and incorporate these various insights to formulate a final response. We explore specifically Retrieval-Augmented Generation (RAG) strategies, both using embedding-based indexing of relevant context and via the use of Ontological Knowledge Graphs that offer a deeper mechanistic delineation of knowledge akin to ontology logs.^{38,54,56–58,61} In such computational approaches, we augment the prompt with relevant information so that the model has access to the expanded context that can include details, measurements, or new data that thereby greatly expand the capabilities of an LLM during generation. We demonstrate these tools against a series of general mechanics/failure knowledge as relevant in a materials design context to show how a fine-tuned LLM combined with Ontological Knowledge Graph strategies can be a particularly powerful combination to expand the use cases of LLMs in scientific

applications that feature both accuracy and interpretability ([Figure 1](#)).

The plan for this paper is to first introduce the general method, show some of the limitations of conventional LLMs, and then move on to the presentation of key ideas and results, agent-based modeling, and various materials design applications, including automatic force-field fitting, followed by a discussion. We conclude with an outlook for future research, especially in the context of materials design.

2. RESULTS AND DISCUSSION

LLMs are powerful tools to answer questions, develop hypotheses, summarize and contrast ideas and content, process complex information, perform tasks, and write codes. In the following sections, we go through a series of computational experiments to probe the performance of LLMs in the various domains, starting with simple retrieval tasks, moving on to graph-based strategies, and ultimately using multiagent AI systems where several models interact via a group chat to solve problems as a team.

2.1. General Query Strategies

We explore first parameter-based knowledge in the fine-tuned LLM, with results for a series of questions shown in [Table 1](#). As can be seen, the model does a reasonable job overall and accurately answers complex domain-specific questions. This is in general agreement with results published in earlier papers.^{28,59} Next, we run the same experiment using Retrieval-Augmented Generation (RAG), with results shown in the right column. RAG is a method where source data from a corpus of information is shared with the LLM during generation, in the general framework of “Considering <INFORMATION CHUNKS>, answer the question <QUESTION>.”. Herein, <INFORMATION CHUNKS> consists of chunks of data retrieved from the corpus of information, according to some measure of relevance for the question asked (this is realized in the work done here using a vector index that identifies relevant data using similarity measures via an embedding vector (for details, see [Materials and Methods](#)). In question answering without RAG, for comparison, we would directly ask the LLM for the answer, akin to “Answer the question <QUESTION>.”. In the first experiment, we observed similar responses. The conclusion of this experiment is that for cases that an LLM can handle on its own, the use of RAG does not significantly improve results.

We now move on to the edge case where the original MechGPT model fails to produce accurate results since we are inquiring about a highly specialized topic. We consider the case of a relatively newly synthesized material, molybdenene, as reported in a recent paper,⁶² as an example. Since this paper was recently published, it was not included in the training data. When we ask MechGPT, the response is (here, the question is in **bold** and the response of the LLM in *italic font*):

What is molybdenene?

Table 1. Comparison of Responses with the Original MechGPT Model and MechGPT with RAG^a

Question	MechGPT	MechGPT with RAG	Notes
Which of these statements is most likely true? (A) Hyperelastic softening leads to higher crack speeds and reduced fracture instabilities, (B) cracks always branch, no matter the hyperelastic effect, (C) Cracks do not exist in any material, (D) hyperelastic softening leads to lower crack speeds and earlier onset of instabilities.	(D) hyperelastic softening leads to lower crack speeds and earlier onset of instabilities.	(D) hyperelastic softening leads to lower crack speeds and earlier onset of instabilities.	Correct answer, for original reference see, e.g. [92]
Considering these deformation mechanisms, which one is most likely to occur during interfacial fracture: (A) supersonic fracture, (B) mother-daughter cracking, (C) grain boundary diffusion, (D) buckling.	Based on the deformation mechanisms discussed, the most likely to occur during interfacial fracture is (B) mother-daughter cracking. This is because mother-daughter cracking involves the propagation of cracks along grain boundaries, which is relevant to the deformation of materials at interfaces.	Based on the information provided, the most likely mechanism to occur during interfacial fracture is (B) mother-daughter cracking.	Correct answer, see references [60,93,94]
Which of these force fields is best suited to describe a novel protein material that serves as synthetic food for fracture studies: (A) ReaxFF, (B) EAM, (C) Morse, or (D) CHARMM.	(A) ReaxFF is the best force field suited to describe a novel protein material that serves as synthetic food for fracture studies.	(D) CHARMM.	Both answers are correct (based on general knowledge about force fields and their use for biomaterials simulations)
Which material is most likely to lead to supersonic fracture in a brittle material: (A) hyperelastic softening, (B) hyperelastic stiffening, (C) linear material behavior.	(B) hyperelastic stiffening is the most likely to lead to supersonic fracture in a brittle material.	(B) hyperelastic stiffening	See, e.g. [95] where it was shown that hyperelastic stiffening can lead to supersonic fracture
Considering these nonlinear material behaviors, which one is most likely to lead to supersonic fracture in a brittle material: (A) hyperelastic softening, (B) hyperelastic stiffening, (C) linear material behavior.	(B) hyperelastic stiffening is most likely to lead to supersonic fracture in a brittle material.	(B) hyperelastic stiffening	Same reference as above
What are the effects of hyperelastic softening on crack speed in brittle materials? Brief answer.	Hyperelastic softening leads to lower crack speeds than expected. It also increases the chances for crack instabilities to set in at lower crack speeds.	Hyperelastic softening leads to lower crack speeds in brittle materials.	See, e.g. [92], where it was shown that softening tends to lead to a reduction of crack speeds in materials
Assuming I have a brittle material that stiffens elastically at large deformation, which of these statements is true: (A) cracks will move supersonically, (B) cracks will move subsonically, (C) cracks will not move at all.	(A) cracks will move supersonically	(A) cracks will move supersonically.	See, e.g. [95] (same explanation as above)
In hyperelastically softening materials, which of these statements is true: (A) cracks will move supersonically, (B) cracks will move at a lower speed than expected, (C) cracks will not move at all.	In hyperelastically softening materials, cracks will move at a lower speed than expected (B).	(B) cracks will move at a lower speed than expected.	See, e.g. [95] (same explanation as above)

Table 1. continued

Question	MechGPT	MechGPT with RAG	Notes
Which of these statements is true? (A) hyperelastic stiffening leads to higher crack speeds and reduced fracture instabilities, (B) cracks always branch, (C) Cracks never move unless heat is applied.	(A) hyperelastic stiffening leads to higher crack speeds and reduced fracture instabilities, (B) cracks always branch, and (C) Cracks never move unless heat is applied are false statements.	(A) hyperelastic stiffening leads to higher crack speeds and reduced fracture instabilities.	See, e.g. [95], where it was shown that hyperelastic stiffening tends to result in higher crack propagation speeds while at the same time displaying a reduced fracture instability tendency (further aiding in achieving high crack speeds)
Which of these statements is most likely true? (A) Hyperelastic stiffening leads to higher crack speeds and reduced fracture instabilities, (B) cracks always branch, no matter the hyperelastic effect, (C) Cracks do not exist in any material.	(A) Hyperelastic stiffening leads to higher crack speeds and reduced fracture instabilities.	(A) Hyperelastic stiffening leads to higher crack speeds and reduced fracture instabilities.	See, e.g. [95] (same explanation as above)
In hyperelastically stiffening materials, which of these statements is true: (A) cracks will move supersonically, (B) cracks will move at a lower speed than expected, (C) cracks will not move at all, (D) cracks will move subsonically.	In hyperelastically stiffening materials, cracks will move supersonically (A).	(A) cracks will move supersonically.	See, e.g. [95] (same explanation as above)
Amongst these materials, which one would most likely show parallel glide dislocations: (A) a silicon wafer, (B) a thin copper film at elevated temperature, (C) a thick copper film at low temperature.	(B) a thin copper film at elevated temperature	(B) a thin copper film at elevated temperature	See, e.g. [96,97] where it was shown that thin copper films show grain boundary diffusion, the formation of a crack-like diffusion wedge, and ultimately the formation of parallel glide dislocations.
Describe the distinction between modeling and simulation.	Modeling refers to the development of a mathematical representation or template of a physical situation, while simulation involves the procedure of solving the equations or numerical problems associated with the model. Simulation yields numbers or trajectories of a system, while modeling provides a framework to understand and predict the behavior of the system.	Modeling involves developing a concept or idea based on physical assumptions and principles, while simulation involves carrying out computations using numerical techniques to solve the resulting partial differential equations. Modeling requires insight into the physics of the system, while simulation requires knowledge in numerical techniques to analyze and interpret the data.	References [60,98] (see text in book and lecture notes)
Which of these mechanisms is most likely to occur in ultra-long carbon nanotubes exposed to compressive loading? (A) chemical etching, (B) buckling, (C) rupture, or (D) supersonic fracture.	Ultra-long carbon nanotubes exposed to compressive loading are most likely to experience (B) buckling. This is because buckling is a common deformation mechanism for long, thin objects like carbon nanotubes, which are weak in compressive strength. Chemical etching (A), rupture (C), and supersonic fracture (D) are less likely scenarios for this specific application.	(B) buckling	See, e.g. [99] where the behavior of ultra-long CNTs were studied, in particular as a function of the aspect ratio (buckling is a dominant mechanism for long nanotubes)

Table 1. continued

Question	MechGPT	MechGPT with RAG	Notes
Discuss how failure in copper is different than failure in a brittle ceramic. Focus on a few key principles, brief answer.	Failure in copper is different than failure in a brittle ceramic. Copper is a metallic material with a high shear modulus and ductility, allowing it to deform plastically before fracturing. Its failure is characterized by dislocation nucleation and propagation in the slip planes. On the other hand, a brittle ceramic like silicon has a high fracture toughness and is prone to brittle fracture. Its failure occurs through the propagation of cracks under tensile stress. The distinction lies in the mode of deformation and fracture, with metals deforming plastically and ceramics fracturing brittlely.	Failure in copper is different than failure in a brittle ceramic. Copper is a ductile metal, meaning it can undergo plastic deformation before breaking. When deformed, dislocations are generated at the crack tip, allowing the material to recover from the applied stress. On the other hand, a brittle ceramic fails by the propagation of cracks, unable to recover the applied stress. The behavior of copper and brittle ceramics are characterized by their atomic bonding structures and the ways in which they deform.	See, e.g. [60,100,101] (brittle materials show crack opening tendencies due to local tensile stresses vs. ductile materials like copper that show dislocations due to local shear on slip planes)
Which of these simulation methods is suitable to describe protein folding: (A) Finite element method, (B) MD simulations, (C) Machine learning/AI.	(B) MD simulations are more suitable to describe protein folding.	(B) MD simulations	See e.g. [102]
Which of these concepts is key to explain the strong adhesion of gecko feet. (A), the use of super-glue by the animal. (B), a size effect, where nano-sized fibrils yield strong and flaw-tolerant adhesion. (C), the low weight of the gecko due to its efficient design.	(B), a size effect, where nano-sized fibrils yield strong and flaw-tolerant adhesion.	(B), a size effect, where nano-sized fibrils yield strong and flaw-tolerant adhesion.	Various sources, including [103]
What is the mirror-mist-hackle instability in dynamic fracture of brittle materials?	The mirror-mist-hackle instability is a phenomenon in dynamic fracture of brittle materials, specifically in the context of atomistic and multiscale modeling. It refers to a crack propagation behavior where the crack surface starts to roughen at around 30% of the Rayleigh wave speed, eventually turning into a hackle region accompanied by the emission of dislocations. Beyond an instability speed, the crack becomes unstable. This behavior can be observed in experiments, computer simulations, and mathematical analysis. The instability is characterized by a roughening of the crack surface and the emission of dislocations, which leads to a transition from a mirror-like crack surface to a hackle region.	The mirror-mist-hackle instability is a dynamic instability in brittle materials, where the crack surface transitions from mirror-like to rough and hackle-filled as the crack propagates. This instability has been observed in LJ systems and is characterized by the emission of dislocations. The final speed of the crack is around 50% of the Rayleigh wave speed.	See e.g. [60] where the phenomenon of mirror-mist-hackle instability was shown for LJ materials (crack surfaces go from mirror-like perfectly flat to misty and increasingly rough, referred to as the hackle regime)
Which statement is correct: (A) Copper can show work-hardening that increases its strength as it is deformed. (B) Copper always fails in a brittle manner. (C) copper only fails in a ductile manner if oxidative species are present.	(A) Copper can show work-hardening that increases its strength as it is deformed.	(A) Copper can show work-hardening that increases its strength as it is deformed.	See, e.g. [101] (and general knowledge about hardening mechanics and metal plasticity)
Is this statement true: Proteins have a greater tendency to unfold under forces applied to the ends of the molecular chain than without force applied. Answer yes or no.	Yes, that's correct. When forces are applied to the ends of a protein molecule, the protein is more likely to unfold compared to when no force is applied. This phenomenon is known as the force-induced unfolding of proteins.	Yes.	Various sources, including [104]

^aAll queries used here are in the domain of materials failure. Since the original model has been well-trained in this area, the responses are accurate throughout. The use of RAG does not significantly improve the answers for this use case since the original MechGPT LLM has already been trained against relevant domain knowledge that is queried in the set of questions.

Molybdenene is a proposed name for the theoretical monolayer of molybdenum atoms. It has not yet been fabricated or observed in experiments.

This response is wrong since this material has actually been fabricated (however, this paper was not included in the training set when the model was developed; it is also noted that sometimes a model may even hallucinate and produce made-up

interfaces, which can play a crucial role in determining the properties of these hybrid materials. Overall, further research and development are needed to explore the full range of possibilities for molybdenene in electronic devices.

Suggest molecular modeling techniques suitable for molybdenene.

Molecular modeling techniques suitable for studying molybdenene include density functional theory (DFT) calculations, molecular dynamics (MD) simulations, Monte Carlo simulations, force-field-based simulations, and *ab initio* calculations. These techniques can provide insights into the molecular structure, electronic properties, mechanical properties, thermal behavior, and interactions of molybdenene with other molecules or materials. Raman spectroscopy can also be used to analyze the vibrational modes of molybdenene, complementing the molecular modeling studies.

Based on your general knowledge and the information provided in the context, what fracture behavior would you expect for molybdenene? Explain why.

Molybdenene is likely to exhibit brittle fracture behavior. This can be inferred from the context information that describes molybdenene as having a layered structure with weak interlayer bonding. The presence of staircase-like features and distinct interatomic distances further suggests a layered structure with weak bonding. A brittle fracture occurs with little or no plastic deformation, and cracks can easily propagate through the material, leading to a sudden failure. The weak interlayer bonding in molybdenene would make it susceptible to brittle fracture as cracks can easily propagate along the layers. Additionally, the observation of local strain-mediated structural evolution and the potential for crack branching and propagation in multiple directions indicate complex fracture behavior, which is often associated with brittle materials. Therefore, based on the information provided, it is reasonable to expect that molybdenene would exhibit brittle fracture behavior.

These answers are generally accurate and agree with the findings reported in the original study.⁶² Specific features about the structure (square lattice, 4-fold symmetry dominance, staircase features) are accurately produced, as can be confirmed by comparing against the original paper, as well as Figure 2c,d (reprinted images from the paper as relevant to the question asked). The model was also asked to suggest relevant modeling tools as well as the anticipated fracture behavior. Key insights from these responses are that they are accurate based on the current understanding. The predicted fracture behavior of “brittle” dominance is explained based on the weak bonding; however, the model also identifies potential complex fracture mechanisms due to deformation-induced structural changes. Such structural changes under mechanical load were indeed identified in the original paper,⁶² and the model has accurately made a connection between this phenomenon and what might happen in a fracture scenario. Of course, more targeted studies of such a scenario are critical, and we should view the LLM response as a direction that points to future hypothesis development that is then followed up by domain experts. In a scenario where an LLM may be used to aid such work, it may be tasked to develop a code or input file for a simulation that can then be executed with results interpreted further. In principle, if relevant tools are implemented, results from such augmented, new data can be fed back into the knowledge basis and used for future generative tasks. This way, the LLM can see more context to make accurate predictions. This workflow can be easily implemented in the current approach.

2.2. Interpretable Graph-Based Strategies

Next, we explore a few other avenues by which the knowledge base is built and queried. In the initial experiment discussed above, we used a vector indexing strategy based on text embeddings to match with a particular query, whereby the Llama Index then assembles proper context for generation using the LLM. Another way to retain functional knowledge is the generation of Ontological Knowledge Graphs, which we find to typically provide better performance since it not only associates embeddings with a query but also develops an ontological understanding of if and how concepts are related. This endows the model with a much deeper context when generating the response and allows it to answer more complex questions.

To start off, we compare responses using conventional RAG and Ontological Knowledge Graphs. The question asked is

Provide a summary of the transcending design principles across different materials classes. Provide examples and details

Here are two exchanges based on ref 63. First, the RAG method based on vector indexing and embeddings:

The transcending design principles across different materials classes involve the assembly of laminated and fibrillar structures, which define a higher level of organization of interfaces and nanostructures. These principles are observed in both bioinspired and synthetic inorganic–organic materials. The combination of strong and weak interfaces determines the toughness, strength, stiffness, shear, and adhesive properties of a material. Examples of materials that exhibit these design principles include bioinspired composites used in drug delivery, wearable electronics, human–computer interfaces, and structural components for the automotive and aerospace industries. These materials possess multifunctional attributes, such as mechanical robustness, flexibility, transparency, sensing, adapting, optics, electrical conductivity, and thermal conductivity. The design principles also involve understanding the biogenesis or “biofabrication” of materials, including the transition from fundamental building blocks to complex hierarchical structures. Digital manufacturing techniques, such as three-dimensional (3D) printing, combined with advances in artificial intelligence, are also being used to create highly complex structures in materials design. The goal is to achieve properties such as optical clarity, loading with bioactive components, and tunable mechanical performance. Overall, the transcending design principles aim to create materials with enhanced properties and functionalities by integrating experiments with theory, modeling, simulation, data science, and artificial intelligence tools.

Second, using Ontological Knowledge Graphs:

The transcending design principles across different materials classes involve the incorporation of hierarchical structures and multiple functionalities. These principles are observed in both natural and synthetic materials. Hierarchical structures play a crucial role in achieving the desired mechanical properties in composite materials. They enable multiple deformation mechanisms, self-healing, plasticity, and toughening across different length scales. At the nanometer scale, compositional gradients and fuzzy interphases facilitate intrinsic toughening through chain slippage, stress delocalization, and nondestructive locking. At the micrometer scale, hierarchical structures contribute to extrinsic toughening through mechanisms such as crack-bridging and energy dissipation via weak or soft interfaces.

Multiple functionalities are achieved by incorporating diverse components and interfaces into the synthetic processes. This involves combining different material phases, creating composition gradients, and enabling reversible energy dissipation. Examples of function-

Table 2. Comparison of Responses of MechGPT Alone and with the Use of RAG for a Specific Topic around Flaw-Tolerance in Graphene^a

Question	MechGPT original	RAG	Notes
Can deep learning models of fracture discover flaw tolerance concepts in graphene? Provide details including critical length scales.	<p>Sure! Here is the answer to the question "Can deep learning models of fracture discover flaw tolerance concepts in graphene? Provide details including critical length scales.":</p> <p>Deep learning models of fracture have been used to study flaw tolerance in graphene. These models have shown that the critical length scale for flaw tolerance in graphene is around 11 nm. This means that below this critical length scale, the stress distribution becomes increasingly homogeneous and eventually uniform near the crack tip. The models have also revealed that the flaw tolerance concept in graphene is closely related to the development of a crack-sealant mechanism. Atomistic modeling of fracture in graphene has further confirmed the importance of the critical length scale in determining flaw tolerance. The study has demonstrated the potential of combining machine learning and atomistic simulations to understand materials failure.</p>	<p>Deep learning models of fracture have the potential to discover flaw tolerance concepts in graphene. These models can capture the complex orientation-dependent fracture behavior of graphene and make predictions about fracture patterns. By analyzing datasets generated by physics-based molecular dynamics simulations, deep learning models can tease out the spatiotemporal relationships underlying fracture propagation.</p> <p>In terms of critical length scales, the deep learning models can provide insights into the fracture behavior of graphene at different orientations. The models have revealed that fracture strength in graphene is reportedly unaffected by flaw sizes below 3.2 nm. However, as the flaw size increases beyond this threshold, significant deviation occurs in the form of additional crack branching. This finding is consistent with the literature, which also reports a threshold of 3.2 nm for nanocrystalline graphene flaw tolerance. It suggests that fracture strength decreases with larger flaw sizes.</p> <p>Overall, deep learning models have demonstrated their ability to uncover flaw tolerance concepts in graphene and provide insights into critical length scales for fracture behavior.</p>	<p>The correct answer is 3.2 nm, see original paper: [105].</p> <p>MechGPT without RAG provides the wrong answer.</p>
Tell me about the fractal dimension of graphene fracture in MD versus ML.	<p>The fractal dimension of graphene fracture in MD (molecular dynamics) versus ML (mechanical loading) is a measure of the complexity and self-similarity of the fracture patterns in graphene. In MD simulations, the fractal dimension is higher, indicating more complex and self-similar fracture patterns. This is compared to ML simulations, where the fractal dimension is lower, indicating more straightforward and less self-similar fracture patterns. The fractal dimension helps in understanding and predicting the behavior of fractures in graphene, which is important for its technological applications.</p>	<p>The fractal dimension of graphene fracture in MD simulations and ML predictions can be compared to evaluate the accuracy of the ML model. The ML-predicted paths tend to underestimate the fractal dimension compared to the real MD results, but the underestimated discrepancy is consistent between most of the trained and test orientations. This suggests that the ML model has a general understanding of graphene fracture behavior. The ML-predicted paths are visually simplified compared to the average MD paths, which may be addressed by increasing the resolution limit of the ML model. Overall, the ML model shows promise in predicting the fractal dimension of graphene fracture.</p>	<p>More details are provided in the RAG method, and answer is correct (MD results show higher fractal dimension, not lower as shown in [64]; MechGPT by itself gets it wrong)</p>
What is original reference that introduced the "box counting" method?	<p>The original reference that introduced the "box counting" method is not mentioned in the provided text.</p>	<p>The original paper that introduced the "box counting" method is Falconer, K. J. The Hausdorff dimension of self-affine fractals. Math. Proc. Camb. Philos. Soc. 103, 339–350 (1988).</p>	<p>Correct reference is provided by RAG, MechGPT does not provide result.</p> <p>Original reference in paper is:</p> <p>37. Falconer, K. J. The Hausdorff dimension of self-affine fractals. Math. Proc. Camb. [106] Philos. Soc. 103, 339–350 (1988).</p>
What is the most salient feature of ML-based fracture modeling?	<p>The most salient feature of ML-based fracture modeling is the use of machine learning algorithms to predict the behavior of materials under different conditions and load modes, leading to faster and more accurate simulations of complex fracture phenomena.</p>	<p>The most salient feature of ML-based fracture modeling is its ability to accurately predict fracture paths in real materials, even in cases with differing surface, grain boundary, and point defects that are different from the original training data.</p>	<p>Response with RAG is more comprehensive and complete and in agreement with key findings reported in [64]. More details are provided.</p>

^aThe RAG method uses the result of a Google Scholar search on the topic of "Deep learning flaw-tolerance", downloading the most highly cited paper (by Lew et al.⁶⁴). Responses include details, facts, correct contexts, and references to the literature.

alities include shape morphing, stimuli-responsiveness, strengthening, toughening, stiffness, chirality, anisotropy, phase transitions, adhesion, self-healing, shear resistance, cracking resistance, and stress redirection.

These design principles and functionalities are observed in various materials classes. For instance, fibrous composites found in nature, such as silk, keratin, and cartilage, exhibit hierarchical structures and multiple functionalities. In engineered heterogeneous materials, ceramic and ballistic armor utilizes Bouligand structures for their



Figure 3. Construction of Ontological Knowledge Graphs from data sources. These graphs are then used by an LLM to provide context to generate responses for queries. Panel a shows an example based on a review paper on natural nanofibrils (reference: ⁶³) and panel b shows an example of a paper on the fracture of intermediate filament networks (ref ⁶⁵). The knowledge graphs shown here describe the entire set of data provided here, scientific papers, that were provided as PDF that is converted to a full-text complement.

optical, acoustic, and mechanical properties. Synthetic bioinspired analogues are being developed to replicate the stiffness, strength, and toughness found in natural materials.

Overall, the transcending design principles involve the incorporation of hierarchical structures and multiple functionalities across different materials classes, both in nature and in synthetic materials. These principles enable materials to achieve the desired mechanical properties and multiple functionalities, making them versatile and adaptable.

The second example is more detailed and includes more mechanistic insights.

As an example to further explore Ontological Knowledge Graphs based generation, we choose a complex domain at the nexus of deep learning and flaw-tolerance that was discussed in the recent literature.⁶⁴ Table 2 depicts a comparison of responses of MechGPT alone and MechGPT and Ontological Knowledge Graphs based RAG for a specific topic around flaw-tolerance in graphene. As can be seen, MechGPT alone fails to provide accurate answers, similar to the earlier example. When we use Ontological Knowledge Graphs built on the top cited result of a Google Scholar search on the topic of “Deep learning flaw-tolerance” (Lew et al.⁶⁴), results are much improved. As can be seen in Table 2, the model now accurately answers detailed questions about this scientific topic. This general approach also allows one to inject new knowledge, or potentially new data/measurements or simulation results, to augment the base baked-in knowledge of the LLM, and it is not limited to scientific reports, papers, patents, and similar but can also incorporate results from physics-based modeling (e.g., *ab initio* or MD or FEM generated data). It demonstrates how the combination of more complex sampling strategies, potentially implemented via

iterative strategies, can provide novel use cases that are particularly relevant for specific scientific scenarios.

We now focus more deeply on graph-based methods and specifically on how they can be utilized effectively for mechanistic insights. Figure 3 depicts details of the process that involves the construction of Ontological Knowledge Graphs from data sources. These graphs are then used to provide context to generate responses for queries. Figure 3a shows an example based on a review paper on natural nanofibrils⁶³ and Figure 3b shows an example of a paper on the fracture of intermediate filament networks based on ref ⁶⁵. A few subgraphs are highlighted in the right panels to reveal the type of insights the model develops from the data provided, clearly showing the similarity in form to what is known as ontology logs.^{56,61,66,67} It is noted that this information is used to provide useful context to a query to help the model understand not only what concepts are important but also how they are related. This additional feature of relatedness, expressed via directional graph edges, is uniquely added to the knowledge graph strategy over a “regular” RAG approach that relies on indexing based on embedding similarities. Specifically, when using graph-based RAG methods, we consider an additional feature of related data chunks derived from the directional graph edges. This means that we do not only collect chunks of relevant context from the vector indexing but also use the knowledge graph to understand how concepts and associated data chunks are related. This is done for up to two hops in the graph structure, thereby infusing concepts and their relationships to one another into the process of question answering. Results for these two specific papers are shown in Tables 3 and 4, respectively. The responses show excellent detail and a lucid tone and provide relevant information.

Table 3. Example of Using Knowledge Graphs to Generate a Response, Which Allows Us to not only Capture Knowledge from Pieces of the Data but also Learn Relationships between Concepts that are Included When the Response is Generated Here for Original Reference by Ling et al., “Nanofibrils in nature and materials engineering”^{63a}

Question	Response	Keywords and knowledge graph context: (knowledge sequence in max depth 2 in the form of directed graph like: 'subject -[predicate]->, object, <- [predicate_next_hop]-, object_next_hop ...')
Tell me about silk nanofibrils.	Silk nanofibrils are biopolymer nanofibrils that possess exceptional mechanical properties. They exhibit a unique combination of high tensile strength and extensibility, surpassing that of steel and Kevlar fibers. Silk nanofibrils have a distinct structure consisting of repetitive core sequences of amino acids with hydrophilic and hydrophobic segments. These segments assemble into random coils and/or helical structures, with the hydrophobic domains transitioning into β -sheet structures. The nanocrystals of silk nanofibrils align along the longitudinal axis, forming fibrils that can further organize into 2D laminar and 3D helicoidal architectures. This helicoidal structure increases the mechanical anisotropy and fracture resistance of the material. It also stiffens the material in the normal direction, enhancing its overall toughness. Silk nanofibrils can be manufactured into different forms such as microspheres, fibers, membranes, gels, foams, scaffolds, and 3D-printed materials. Due to their remarkable mechanical performance, sustainability, availability, low-cost production, low weight, and biocompatibility, silk nanofibrils have great potential for various large-scale applications. They have been utilized in numerous applications, including drug and cell carriers, mechanically enhanced materials, flexible optic and electronic devices, water filtration, and improving the mechanical properties of structural materials.	Extracted keywords: ['nanofibrils', 'silk'] (('silk nanofibrils', 'represent', 'three of the most abundant biopolymer nanofibrils') ['nanofibrils', 'are good candidates', 'because of absence of physical entanglements'] (('silk nanofibrils', 'possess', 'unparalleled mechanical properties') ['silk', 'can be spun into', 'oriented fibres'] ['nanofibrils', 'aligned as', 'planar layers']
How can I build a new filter selective for virus particles using silk nanofibrils	To build a new filter selective for virus particles using silk nanofibrils, one approach is to fabricate specific periodic shapes in the silk nanofibrils. This can be achieved through techniques such as microfluidic spinning, which regulates the assembly of biopolymer nanofibrils. By controlling the shape and arrangement of the nanofibrils, the filter can be designed to effectively capture and filter out virus particles while allowing other substances to pass through. Additionally, incorporating twisted layers of silk nanofibrils in the filter can enhance its mechanical anisotropy and improve fracture resistance, similar to the helicoidal structure observed in chitin nanofibrils found in lobster cuticles. The hierarchical structure and mechanical properties of silk nanofibrils can be leveraged to optimize the performance of the filter in selectively capturing virus particles.	Extracted keywords: ['build', 'new filter', 'silk nanofibrils', 'virus', 'new', 'nanofibrils', 'silk', 'filter', 'virus particles', 'particles', 'selective'] (('Optically anisotropic silk nanofibrils', 'can be generated by', 'fabricating specific periodic shapes') ['nanofibrils', 'are good candidates', 'because of absence of physical entanglements'] ['silk nanofibrils', 'represent', 'three of the most abundant biopolymer nanofibrils'] (('microfluidic spinning', 'regulate', 'biopolymer nanofibril assembly') ['silk nanofibrils', 'possess', 'unparalleled mechanical properties'] ['silk nanofibrils', 'mixed with', 'hydroxyapatite'] ['silk', 'can be spun into', 'oriented fibres'] ['nanofibrils', 'aligned as', 'planar layers']

^aTo do this, a PDF of the paper is analyzed using an LLM, and an Ontological Knowledge Graph is constructed using GPT-3.5-turbo (the graph can be seen in Figure 3a). The knowledge graph is then used to construct responses using MechGPT. Such a graph is a way of visualizing data across different representations and is composed of three main components: Nodes, edges, and labels. Any concept, material, method, person, etc., can be a node. An edge defines the relationship between the nodes.

We now move to explore interpretability for specific tasks executed by the LLM, not the entire knowledge graph. Indeed, we can directly visualize the task-specific Ontological Knowledge Graphs, here exemplified for the deformation behavior of protein filament networks (Figure 4a, as reported in ref 65). For instance, Figure 4b shows a knowledge graph for the question:

Tell me more about flaw-tolerance in protein networks.

For the case of the intermediate filament paper 65. The extracted keywords include “tolerance”, “flaw”, “protein”, “networks”, “flaw-tolerance”, and “protein networks”. From this, a task-specific Ontological Knowledge Graph is constructed and visualized in Figure 4b, with additional context provided in Table 5. The model response is:

Flaw-tolerance in protein networks refers to their ability to withstand and tolerate structural irregularities or flaws without catastrophic failure. Unlike conventional solids, protein networks, particularly α -helical protein networks, can dissipate a significant amount of energy and stretch without affecting neighboring bonds. This is because there are no immediate interactions between individual filaments in the network that prevent microscopic rotations and shear, resulting in no strong stress concentrations at the corners of cracks. The flaw-tolerance of protein networks is attributed to their unique geometric arrangement, the properties of individual protein domains, and the ability to dissipate energy and stretch without affecting neighboring bonds. Additionally, protein networks have a self-healing nature, as the utilization of hydrogen

Table 4. Based on the Scientific Paper Titled “ α -Helical Protein Networks are Self-Protective and Flaw-Tolerant” by Ackbarow et al.,⁶⁵ We Develop an RAG-Based Set of Model Queries and Responses (the Graph can be Seen in Figure 3b)^a

Question	Response	Keywords and knowledge graph context: (knowledge sequence in max depth 2 in the form of a directed graph like: `subject -[predicate]->, object, <- [predicate_next_hop]-, object_next_hop ...`)
What is the origin of flaw-tolerance in alpha-helical protein networks?	The origin of flaw-tolerance in alpha-helical protein networks is attributed to the unique nanomechanical properties of the protein constituents. These properties enable the formation of extensive dissipative yield regions around structural flaws, providing effective protection against catastrophic failure. Additionally, a geometric transformation of the crack shape helps minimize stress concentration at corners, further enhancing the flaw-tolerance of alpha-helical protein networks.	<p>Extracted keywords: ['protein', 'origin', 'flaw', 'networks', 'flaw-tolerance', 'helical', 'tolerance', 'alpha', 'alpha-helical protein networks']</p> <p>['alpha-helical protein networks', 'largely insensitive', 'to structural flaws'] ['alpha-helical protein networks', 'is published in', 'PLoS ONE'] ['alpha-helical protein networks', 'are', 'first of its kind'] ['networks', 'do not display', 'strong stress concentration'] ('Alpha-helix based protein networks', 'withstand', 'flaws') ('Alpha-helical protein networks', 'are', 'flaw-tolerant') ['PLoS ONE', 'is published in', 'June 2009'] ['networks', 'break at', 'low strains'] ['networks', 'break at', 'low stress'] ['PLoS ONE', 'is', 'www.plosone.org'] ['PLoS ONE', 'is', 'June 2009'] ['PLoS ONE', 'is', 'Volume 4'] ['PLoS ONE', 'is', 'Issue 6'] ['PLoS ONE', 'is', 'e6015']</p>
Tell me more about the geometric transformation of the crack shape helps minimize stress concentration at corners.	The geometric transformation of the crack shape is believed to help minimize stress concentration at corners, although the exact mechanism is not fully understood. As the crack undergoes a series of changes, from mode I to a circular hole, and eventually becoming an elongated crack aligned with the direction of loading, the stress concentration at corners is reduced. This transformation occurs as individual proteins within the crack continuously unfold, independent of their neighboring proteins. The reduced stresses in the vicinity of the elongated crack contribute to a decrease in stress concentration at corners. Overall, this geometric transformation of the crack shape is thought to provide self-protection and distribute stress more evenly along the crack, ultimately minimizing stress concentration.	<p>Extracted keywords: ['transformation', 'corners', 'minimize', 'concentration', 'crack', 'stress', 'crack shape', 'stress concentration', 'geometric', 'geometric transformation', 'shape']</p> <p>The following are knowledge sequence in max depth 2 in the form of directed graph like: `subject -[predicate]->, object, <-[predicate_next_hop]-, object_next_hop ...` ['stress concentration', 'can be seen by', 'red color indicating stretching of the protein filament's covalent backbone'] ['stress', 'drop to zero', 'once this failure mode begins to operate'] ['geometric transformation', 'described by', 'j0-C1j and d0-C2d'] ('Crack shape', 'reduces', 'stress concentration at corners') ('stress concentrations', 'corner of', 'cracks') ['crack shape', 'changes', 'under applied load'] ['transformation', 'of', 'crack shape'] ['crack', 'propagates', 'leading to'] ['strength', 'is', '<600 MPa'] ['crack', 'has', 'strength']</p>

Table 4. continued

Question	Response	Keywords and knowledge graph context: (knowledge sequence in max depth 2 in the form of a directed graph like: `subject -[predicate]->, object, <- [predicate_next_hop]-, object_next_hop ...`)
<p>"The mechanism of flaw-tolerance in alpha-helical protein networks is: The geometric transformation of the crack shape is believed to help minimize stress concentration at corners, although the exact mechanism is not fully understood. As the crack undergoes a series of changes, from mode I to a circular hole, and eventually becoming an elongated crack aligned with the direction of loading, the stress concentration at corners is reduced. This transformation occurs as individual proteins within the crack continuously unfold, independent of their neighboring proteins. The reduced stresses in the vicinity of the elongated crack contribute to a decrease in stress concentration at corners. Overall, this geometric transformation of the crack shape is thought to provide self-protection and distribute stress more evenly along the crack, ultimately minimizing stress concentration. Now, describe how to translate this concept to a material composed out of thin glass fibers as elementary building blocks. Define details of how the glass fibers must be engineered, and how they are arranged into larger-scale structures."</p>	<p>To achieve flaw-tolerance in a material composed of thin glass fibers as elementary building blocks, it is important to engineer the glass fibers in a way that allows for geometric transformations similar to those observed in alpha-helical protein networks. This involves designing the glass fibers with specific properties that enable them to undergo unfolding and reformation of bonds under stress, similar to the unfolding of individual proteins in alpha-helical protein networks.</p> <p>In addition to the engineering of the glass fibers, the arrangement of these fibers into larger-scale structures is crucial. The glass fibers should be arranged in a manner that allows for independent movement and deformation of individual fibers, similar to the arrangement in alpha-helical protein networks. This can be achieved by creating a network or lattice-like structure where the fibers are interconnected but still able to move independently.</p> <p>By engineering the glass fibers and arranging them in a suitable structure, it is possible to achieve flaw-tolerance similar to that observed in alpha-helical protein networks. This involves designing the glass fibers to undergo geometric transformations that help minimize stress concentration at corners, ultimately distributing stress more evenly along the crack. The specific details of how the glass fibers must be engineered and arranged into larger-scale structures would depend on the properties of the glass material and the desired level of flaw-tolerance.</p>	<p>Extracted keywords: ['corners', 'individual', 'thin glass fibers', 'self', 'individual proteins', 'networks', 'flaw-tolerance', 'protection', 'elementary building blocks', 'reduced', 'distribute stress', 'vicinity', 'loading', 'unfold', 'stresses', 'elongated', 'protein', 'flaw', 'direction', 'geometric', 'proteins', 'mechanism', 'circular hole', 'thin', 'larger', 'reduced stresses', 'larger-scale structures.', 'glass', 'helical', 'alpha', 'self-protection', 'engineered', 'arranged', 'concentration', 'crack', 'blocks', 'mode I', 'tolerance', 'I', 'building', 'geometric transformation', 'shape', 'transformation', 'scale', 'distribute', 'hole', 'direction of loading', 'mode', 'alpha-helical protein networks', 'structures', 'fibers', 'elementary', 'stress', 'crack shape', 'circular', 'stress concentration', 'elongated crack']</p> <p>('there is a slight stress concentration at the tip of the crack', 'can be seen by', 'red color indicating stretching of the protein filament's covalent backbone') ['stress concentration', 'can be seen by', 'red color indicating stretching of the protein filament's covalent backbone'] ['alpha-helical protein networks', 'largely insensitive', 'to structural flaws'] ['stress', 'drop to zero', 'once this failure mode begins to operate'] ('alpha-helical protein network', 'under the presence of', 'crack') ['geometric transformation', 'described by', 'j0-C1j and d0-C2d'] ['alpha-helical protein networks', 'is published in', 'PLOS ONE'] ['alpha-helical protein networks', 'are', 'first of its kind'] ['networks', 'do not display', 'strong stress concentration'] ['elongated crack', 'features', 'small stresses'] ['crack shape', 'changes', 'under applied load'] ['PLOS ONE', 'is published in', 'June 2009'] ['circular hole', 'to', 'elongated crack'] ['transformation', 'of', 'crack shape'] ['networks', 'break at', 'low strains'] ['networks', 'break at', 'low stress'] ['PLOS ONE', 'is', 'www.plosone.org'] ['crack', 'propagates', 'leading to'] ['PLOS ONE', 'is', 'June 2009'] ['PLOS ONE', 'is', 'Volume 4'] ['strength', 'is', '<600 MPa'] ['PLOS ONE', 'is', 'Issue 6'] ['crack', 'has', 'strength'] ['PLOS ONE', 'is', 'e6015']</p>

⁴For comparison, a query of the MechGPT without RAG gives "The origin of flaw-tolerance in α -helical protein networks is believed to be the ability of the network to redistribute stress and strain around flaws or defects. This redistribution helps prevent the propagation of cracks or fractures, thereby enhancing the overall mechanical integrity of the network." While this answer is generally correct, as can be seen in the examples below, graph knowledge-based extraction provides more detailed, mechanistic insights.

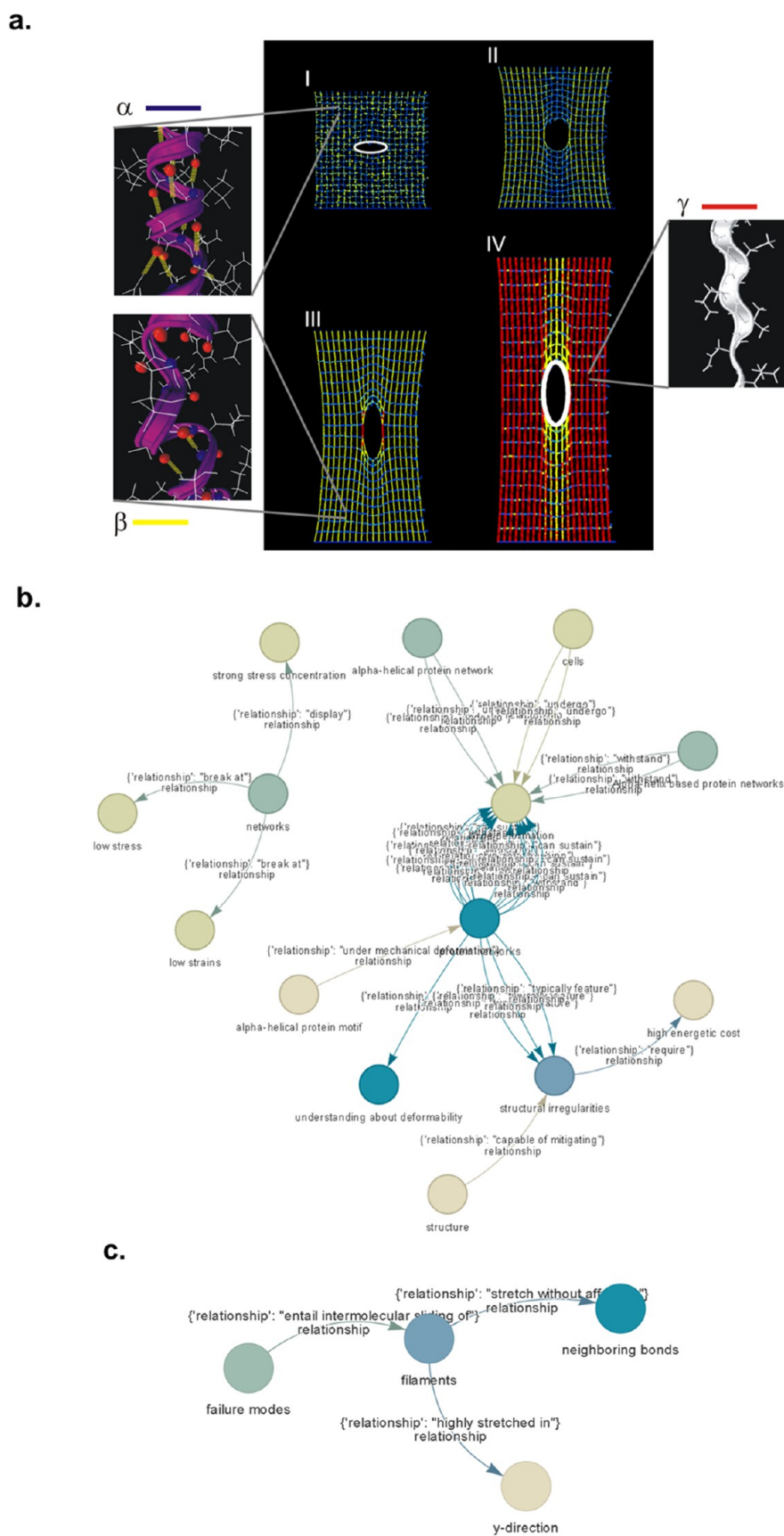


Figure 4. Construction and visualization of an Ontological Knowledge Graph that is constructed and used to answer a question about a protein-based filament network⁶⁵ (panel a showing the deformation behavior and key mechanisms leading to flaw-tolerance via a structural transformation of the crack, mediated by the underlying mesh structure and protein filament stretching via helix unfolding). The query used is “Tell me more about flaw-tolerance in protein networks.” (panel b) and “What is the mechanism by which filaments ultimately fail?” (panel c). The graphs shown here are specific to

Figure 4. continued

the question asked and explain a graphical visualization by which the task is addressed. This is in contrast to the graphs shown in Figure 3, which depict the entire ontological knowledge of a larger data set. Images in panel a reprinted or adapted with permission under a Creative Commons CC BY 4.0 license from ref 65. Copyright 2009 Public Library of Science (PLOS).

Table 5. Raw Definition of the Ontological Knowledge Graphs Constructed for the Query “Tell Me More About Flaw-Tolerance in Protein Networks”, Obtained Using NebulaGraph, which Provides Detailed Information on the Relationship between Nodes via Edges

networks{name: networks} -[relationship:{relationship: display}]> strong stress concentration{name: strong stress concentration}
networks{name: networks} -[relationship:{relationship: break at}]> low strains{name: low strains}
networks{name: networks} -[relationship:{relationship: break at}]> low stress{name: low stress}
protein networks{name: protein networks} -[relationship:{relationship: typically feature}]> structural irregularities{name: structural irregularities} -[relationship:{relationship: require}]> high energetic cost{name: high energetic cost}
protein networks{name: protein networks} -[relationship:{relationship: can sustain}]> large deformation{name: large deformation} <-[relationship:{relationship: undergo}]- cells{name: cells}
protein networks{name: protein networks} -[relationship:{relationship: typically feature}]> structural irregularities{name: structural irregularities}
protein networks{name: protein networks} -[relationship:{relationship: can sustain}]> large deformation{name: large deformation} <-[relationship:{relationship: undergo}]- alpha-helical protein network{name: alpha-helical protein network}
protein networks{name: protein networks} -[relationship:{relationship: can sustain}]> large deformation{name: large deformation} <-[relationship:{relationship: withstand}]- Alpha-helix based protein networks{name: Alpha-helix based protein networks}
protein networks{name: protein networks} -[relationship:{relationship: withstand}]> large deformation{name: large deformation} <-[relationship:{relationship: withstand}]- Alpha-helix based protein networks{name: Alpha-helix based protein networks}
protein networks{name: protein networks} -[relationship:{relationship: withstand}]> large deformation{name: large deformation}
protein networks{name: protein networks} -[relationship:{relationship: typically feature}]> structural irregularities{name: structural irregularities} <-[relationship:{relationship: capable of mitigating}]- structure{name: structure}
protein networks{name: protein networks} -[relationship:{relationship: withstand}]> large deformation{name: large deformation} <-[relationship:{relationship: undergo}]- cells{name: cells}
protein networks{name: protein networks} -[relationship:{relationship: advance}]> understanding about deformability{name: understanding about deformability}
protein networks{name: protein networks} <-[relationship:{relationship: under mechanical deformation}]- alpha-helical protein motif{name: alpha-helical protein motif}
protein networks{name: protein networks} -[relationship:{relationship: withstand}]> large deformation{name: large deformation} <-[relationship:{relationship: undergo}]- alpha-helical protein network{name: alpha-helical protein network}
protein networks{name: protein networks} -[relationship:{relationship: can sustain}]> large deformation{name: large deformation}
protein networks{name: protein networks} -[relationship:{relationship: withstand}]> large deformation{name: large deformation} <-[relationship:{relationship: can sustain}]- protein networks{name: protein networks}
protein networks{name: protein networks} -[relationship:{relationship: can sustain}]> large deformation{name: large deformation} <-[relationship:{relationship: withstand}]- protein networks{name: protein networks}

bonds allows the structure to reform at moderate temperatures and restore the initial α -helical structure even after severe deformation. This flaw-tolerance allows protein networks to withstand large deformations and stresses, making them highly resilient and mechanically robust. Understanding the mechanisms behind the flaw-tolerance of protein networks is important for developing models that describe the structure–property relationships in these networks under large deformation and for advancing knowledge in fields such as materials science and bioengineering.

We now proceed to the response to the question:

What is the mechanism by which filaments ultimately fail?

Extracted keywords are “filaments”, “fail”, and “mechanism”.

The response is

The mechanism by which filaments ultimately fail is through the stretching of the α -helical protein domains, which leads to the

breaking of H-bonds within and between the filaments. This stretching and breaking of bonds eventually cause the filaments to fail.

The corresponding Ontological Knowledge Graphs are shown in Figure 4c. The graphs in Figure 4b,c provide detailed mechanistic insights into how the model recalls knowledge and answers the questions. A close inspection of these graphs can provide additional insights into the mechanisms of a system of interest and thereby provide not only interpretability but also the potential to develop new hypotheses or research questions that can be realized in an AI-human collaboration.

Another is done using more sophisticated sampling methods that build on the concept of adapting the way we interact with an LLM and introducing alternatives to simple linear sampling, using techniques such as Chain-of-Thought prompting,¹² Tree-

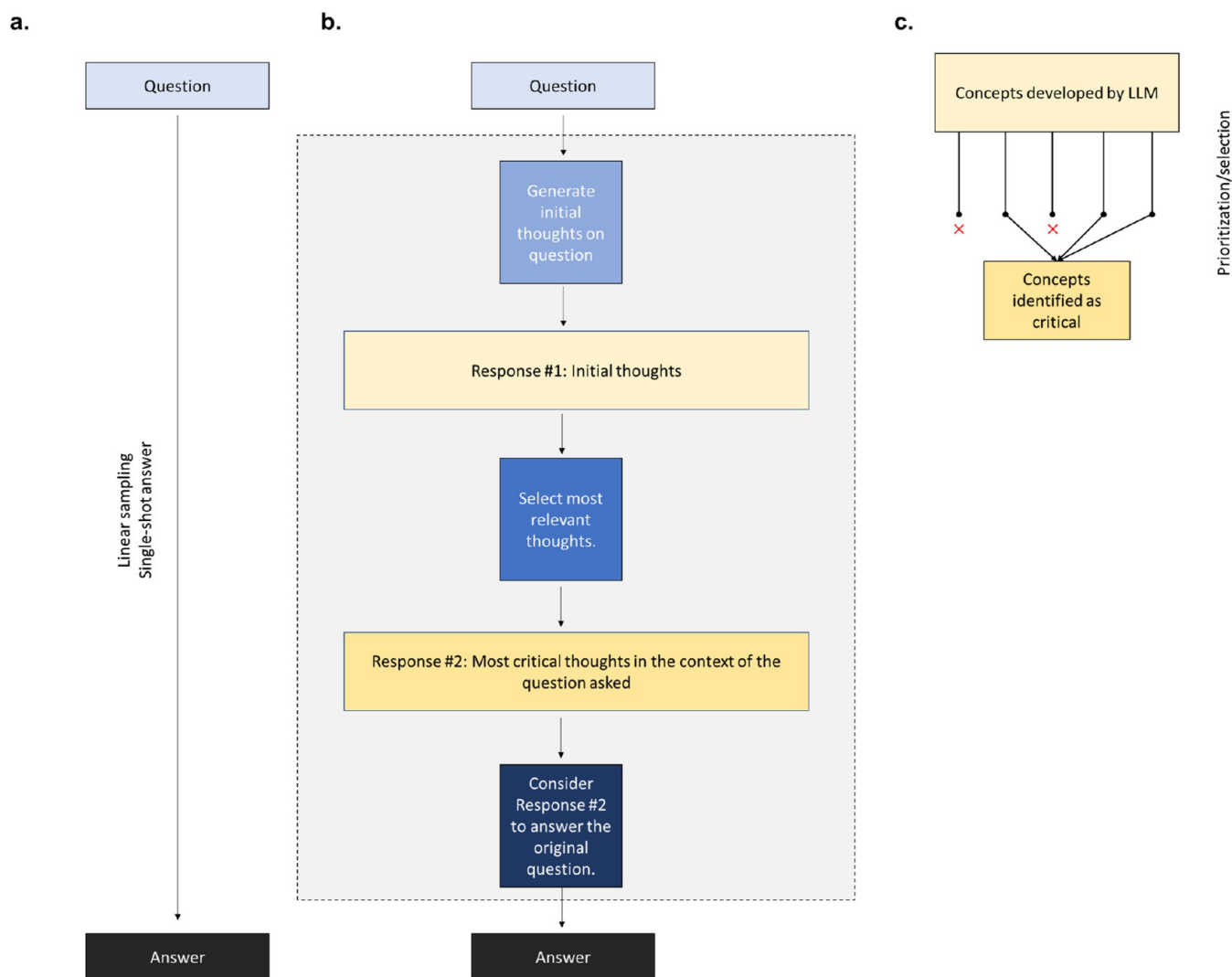


Figure 5. Construction of responses based on direct, linear sampling (panel a) versus hierarchical tree sampling (panel b). Here, the LLM is first used to generate initial thoughts. Then, in the second step, the model is queried to develop the most critical concepts and list them. This set of critical concepts is then used to answer the original question, to yield the answer. This stepwise interactive use of LLMs can provide advantages to extract more accurate, more nuanced, and more refined knowledge. As visualized in panel c, the assessment of responses, such as concepts using an LLM serving the role of a critic, can provide powerful nonlinear sampling mechanisms that significantly enhance responses.

of-Thought strategies,⁶⁸ and related approaches.⁶⁹ Figure 5 shows the basic mechanics of this approach, whereas Figure 5a shows the construction of responses based on direct, linear sampling versus more complex, multiple-step sampling (Figure 5b). As depicted in Figure 5c, the assessment of responses, such as concepts using an LLM serving the role of a critic, can provide powerful nonlinear sampling mechanisms that significantly enhance responses (and a multitude of ways can be implemented depending on use cases). In its simplest form, the general idea is that the LLM is first used to generate initial thoughts (and we can enhance the creativity of the model by increasing the sampling temperature—this may result in more explorative answers). Then, in the second step, the model is queried to develop the most critical concepts and list them (these steps can be repeated multiple times). This set of critical concepts is then used to answer the original question, to yield the answer. This stepwise interactive use of LLMs can provide advantages to extract more accurate, more nuanced, and more refined knowledge.

Table 6 shows a specific example using multistep tree sampling. The question asked is

What would be the likely failure mechanisms of a hybrid silk–metal nanocomposite reinforced with carbon nanotubes?

We proceeded by asking several questions to sample dynamically. Compared with the single-shot answer, the complex prompting strategy elicits more nuanced and detailed responses. Table 7 shows another result using multistep tree sampling. Here, the question asked is

The mechanism of flaw-tolerance in α -helical protein networks is “The geometric transformation of the crack shape is believed to help minimize stress concentration at corners, although the exact mechanism is not fully understood. As the crack undergoes a series of changes, from mode I to a circular hole and eventually becoming an elongated crack aligned with the direction of loading, the stress concentration at corners is reduced. This transformation occurs as individual proteins within the crack continuously unfold, independent of their neighboring proteins. The

Table 6. Multistep Sampling Inspired by the Tree-of-Thought Method^a

Question	Response	Notes
Generate a list of initial thoughts that are relevant for answering this question: '{question}'. Do not answer the question.	{Response #1}= 1. Fracture of the silk fibers due to tensile strain 2. Delamination between the silk and metal layers 3. Debonding between the silk and carbon nanotubes 4. Fracture of the carbon nanotubes 5. Deformation of the metal components These failure mechanisms consider the different materials present in the hybrid nanocomposite and take into account their individual properties and interactions.	
Read this: '{Response #1}'. List the most important concepts to answer the question '{question}'.	{Response #2}= The most important concepts to answer the question 'What would be likely failure mechanisms of a hybrid silk-metal nanocomposite, reinforced with carbon nanotubes?' are fracture of the silk fibers due to tensile strain, delamination between the silk and metal layers, debonding between the silk and carbon nanotubes, fracture of the carbon nanotubes, and deformation of the metal components.	Here we use the LLM to identify the most critical concepts to answer the question.
Considering '{Response #2}', answer this question with a detailed response: '{question}'	The likely failure mechanisms of a hybrid silk-metal nanocomposite, reinforced with carbon nanotubes, include fracture of the silk fibers due to tensile strain, delamination between the silk and metal layers, debonding between the silk and carbon nanotubes, fracture of the carbon nanotubes, and deformation of the metal components.	
Single-shot answer to the question.	The likely failure mechanisms of a hybrid silk-metal nanocomposite, reinforced with carbon nanotubes, would be determined by the interplay between the silk and metal components. The silk fibers, being natural materials, are susceptible to biodegradation and environmental factors. The metal nanoparticles, on the other hand, provide strength and durability but may also be prone to corrosion and wear. The carbon nanotubes, being carbon-based, can add stability and conductivity to the material but may also be affected by environmental factors. Overall, the failure mechanisms would depend on the specific conditions and interactions between the silk, metal, and carbon nanotubes.	Answer is more general, lacks details on mechanisms.

^aThe question asked is “What would be likely failure mechanisms of a hybrid silk–metal nanocomposite, reinforced with carbon nanotubes?”. We proceed by asking several questions to sample dynamically. Compared to the single-shot answer, the complex prompting strategy elicits more nuanced and detailed responses that generally agree with the literature of related materials.¹⁰⁷

reduced stresses in the vicinity of the elongated crack contribute to a decrease in the stress concentration at corners. Overall, this geometric transformation of the crack shape is thought to provide self-protection and distribute stress more evenly along the crack, ultimately minimizing stress concentration.” Now, we describe how to translate this concept to a material composed of thin glass fibers as elementary building blocks. Define details of how the glass fibers must be engineered and how they are arranged into larger-scale structures.

We then proceeded by asking several follow-up questions, with detailed results of the experiment provided via Table 7. Compared to the single-shot answer, the complex prompting strategy elicits more nuanced and detailed responses. This example showed how additional augmentation and context can be built into the query itself to then trigger a rich set of generative steps with lots of nuanced responses that are helpful for engineering and scientific processes.

Finally, the opportunities of using LLMs offer interesting possibilities at the interface of knowledge domains, especially using a variety of Ontological Knowledge Graphs. Figure 6 shows the results of such an experiment, where we use MechGPT to elucidate analogies between disparate domains, in this case, between music and proteins. We use the MechGPT model to sample a set of responses to the query:

Discuss the analogies between music and proteins. Specifically, address counterpoint and how it relates music to protein. Provide a detailed and lengthy answer.

Using GPT-4, we then assemble an Ontological Knowledge Graph from this set of data to elucidate key connections and relationships identified. We find that the key connectors between the various subgraphs are “basic units”, where these are amino acids in proteins and notes in music. The model further predicts several specific analogy concepts, such as that

analogies “lie in the organization of units” and “help us understand wonders of life”, among many others (subgraph in the right lower corner). Other connectors between subgraphs are identified as “structures” and “notes” as key nodes. This example can be expanded in future work, but it shows the capacity of using LLMs as a design tool to connect disparate domains of knowledge and to interact with data and knowledge in innovative ways that can augment conventional engineering workflows.

2.3. Multi-AI Agent-Based Strategies: Solving Tasks Using a Team of LLMs

The previous examples were conducted by using one LLM to solve tasks ranging from question answering to graph generation. We now show how a set of LLM agents, organized into a virtual team of AIs (Figure 1d), can effectively solve complex problems via collaborative interactions. The individual agents can have distinct properties, knowledge, and special powers. For instance, in the first example, we create agents that have special powers to generate molecular coordinates from SMILES⁷⁰ representations of arbitrary molecules and also have powers to conduct a Density Function Theory (DFT) simulation.⁷¹ The ability to conduct DFT simulations, in particular, is highly relevant since it provides access to new physics-based data; here, agents have the ability to complement baked-in knowledge from pretraining with new measurements or simulation results.

2.3.1. Molecular Design. As a first example, Table 8 shows a result of such agent-based modeling applied here to solve a molecular design problem. The question posed here is “Consider this molecule in SMILES representation: CCCCXCC, where X is one of (C, O, N). Which of these options leads to the lowest energy structure?” The agents devise a plan, carry out the simulations, analyze the data, and provide the final response to the user. The

Table 7. Multistep Sampling Inspired by the Tree-of-Thought Method, Implemented via Variegated Uses of the LLM as a Generator of Ideas/Concepts and to Judge These against Their Usefulness to Solve a Particular Task^a

Question	Response	Notes
Generate a list of initial thoughts that are relevant for answering this question: '{question}'. Do not answer the question.	{Response #1}= Sure, here are some initial thoughts relevant to answering the question: - The mechanism of flaw-tolerance in alpha-helical protein networks is believed to involve the geometric transformation of the crack shape, which helps minimize stress concentration at corners. - As the crack undergoes changes, the stress concentration at corners is reduced, and the stress is distributed more evenly along the crack. - Individual proteins within the crack unfold independently, contributing to the reduction of stress concentration. - The overall effect is a self-protection mechanism that distributes stress more evenly, minimizing stress concentration. To apply this concept to a material composed out of thin glass fibers as elementary building blocks, the glass fibers must be engineered to have specific properties. The arrangement of the glass fibers into larger-scale structures is crucial to ensure the desired strength and flaw-tolerance. The exact engineering details may need to be determined based on the specific requirements of the material.	In the first step we ask MechGPT to generate a set of initial concepts relevant to the query.
Read this: '{Response #1}'. List the most important concepts to answer the question '{question}'.	{Response #2}= The mechanism of flaw-tolerance in alpha-helical protein networks is believed to involve the geometric transformation of the crack shape, which helps minimize stress concentration at corners. As the crack undergoes changes, the stress concentration at corners is reduced, and the stress is distributed more evenly along the crack. Individual proteins within the crack unfold independently, contributing to the reduction of stress concentration. The overall effect is a self-protection mechanism that distributes stress more evenly, minimizing stress concentration. To apply this concept to a material composed out of thin glass fibers as elementary building blocks, the glass fibers must be engineered to have specific properties. The arrangement of the glass fibers into larger-scale structures is crucial to ensure the desired strength and flaw-tolerance. The exact engineering details may need to be determined based on the specific requirements of the material.	Here we use the LLM to identify the most critical concepts to answer the question.
Considering '{Response #2}', answer this question with a detailed response: '{question}'	To apply this concept to a material composed out of thin glass fibers as elementary building blocks, the glass fibers must be engineered to have specific properties. The arrangement of the glass fibers into larger-scale structures is crucial to ensure the desired strength and flaw-tolerance. The exact engineering details may need to be determined based on the specific requirements of the material.	Now we prompt the LLM to answer the question. Key issues reproduced agree with our understanding of the key aspects, including larger-scale mesh structures, and the significance of ensuring that the fibers themselves have specific properties (akin to the protein filaments that can stretch exhaustively without breaking, as laid out in [65,108]).
Single-shot answer to the question.	Sure! What is the question regarding the engineering of glass fibers as elementary building blocks in a material?	No answer provided. Alternative responses may include hallucinated results.

^aThe question asked is “The mechanism of flaw-tolerance in α -helical protein networks is “The geometric transformation of the crack shape is believed to help minimize stress concentration at corners, although the exact mechanism is not fully understood. As the crack undergoes a series of changes, from mode I to a circular hole, and eventually becoming an elongated crack aligned with the direction of loading, the stress concentration at corners is reduced. This transformation occurs as individual proteins within the crack continuously unfold, independent of their neighboring proteins. The reduced stresses in the vicinity of the elongated crack contribute to a decrease in stress concentration at corners. Overall, this geometric transformation of the crack shape is thought to provide self-protection and distribute stress more evenly along the crack, ultimately minimizing stress concentration.” Now, describe how to translate this concept to a material composed of thin glass fibers as elementary building blocks. Define details of how the glass fibers must be engineered and how they are arranged into larger-scale structures.” We proceed by asking several questions to sample dynamically and iteratively by using MechGPT to judge and assess its own responses. Compared to the single-shot answer, the complex prompting strategy elicits more nuanced and detailed responses. It is noted that in this example, we do not use RAG, and only the question designated above is used to prompt the model.

conversational transcript shown in the table provides a detailed account of how the problem is solved and how the answer is ultimately provided. The set of agents includes the following entities:

- “User” (human who poses the question),

- “Planner” who develops a plan to answer the question,
- “Coordinate retriever” that can retrieve coordinates from SMILES coordinates or its own knowledge
- “Chatbot” with broad skills and execution power.

Table 8. Agent-Based Modeling, Applied Here to Solve a Molecular Design Problem⁴²

Conversation	Notes
<p>User: Consider this molecule in SMILES representation: CCCCXCC, where X is one of (C, O, N). Which of these options leads to the lowest energy structure?</p>	Initial question posed
<p>Planner: Sure, to solve this task we need to convert the SMILES representation into coordinates for each molecule, calculate the energy for each molecule, and then compare the energy level.</p> <p>Let's detail this in a precise plan:</p> <ol style="list-style-type: none"> For each of the atoms C, O, N as X: <ul style="list-style-type: none"> Replace X in the SMILES "CCCCXCC" with the atom. Use the function 'functions.coords_from_SMILES' with the new SMILES representation to get the coordinates of the molecule. Use the function 'functions.query_DFT' with these coordinates to calculate the energy of the molecule. Compare the calculated energy levels for the three molecules. The atom that results in the lowest energy level molecule is the one that should replace X. 	Planner develops a strategy to solve the problem
<p>Coordinate retriever: Understood. Let's get started with step 1, where we'll generate coordinates for each of the molecules and calculate their energies.</p> <p>***** Suggested function Call: coords_from_SMILES *****</p> <p>Arguments:</p> <pre>{ "SMILES": "CCCCCCC" }</pre>	The other agents follow the plan and retrieve information (starting with coordinate retrieval for each of the structures, energy calculation, etc.); repeated for each of the three cases of X=(C, O, N).
<p>>>>>>>> EXECUTING FUNCTION coords_from_SMILES...</p> <p>Chatbot:</p> <p>***** Response from calling function "coords_from_SMILES" *****</p> <pre>[C 2.459 0.3911 0.8023, 'C 2.528 -0.475 -0.4535', 'C 1.304 -0.309 -1.369', 'C -0.006061 -0.8418 -0.7517', 'C -0.8906 0.2868 -0.2008', 'C -2.163 -0.2759 0.4426', 'C -3.05 0.8428 0.9815', 'H 1.67 0.01777 1.488', 'H 2.254 1.45 0.5352', 'H 3.43 0.3431 1.338', 'H 2.644 -1.542 -0.1625', 'H 3.436 -0.1851 -1.027', 'H 1.203 0.7551 -1.677', 'H 1.508 -0.8856 -2.298', 'H 0.2104 -1.583 0.04887', 'H -0.5783 -1.375 -1.543', 'H -1.172 0.9684 -1.034', 'H -0.3301 0.8759 0.5549', 'H -1.891 -0.9536 1.281', 'H - 2.738 -0.8594 -0.3097', 'H -3.962 0.4059 1.441', 'H -2.507 1.426 1.755', 'H -3.358 1.522 0.158'] *****</pre> <p>***** Suggested function Call: query_DFT *****</p> <p>Arguments:</p> <pre>{ "coordinates": "C 2.459 0.3911 0.8023\nC 2.528 -0.475 -0.4535\nC 1.304 -0.309 -1.369\nC -0.006061 -0.8418 -0.7517\nC -0.8906 0.2868 - 0.2008\nC -2.163 -0.2759 0.4426\nC -3.05 0.8428 0.9815\nH 1.67 0.01777 1.488\nH 2.254 1.45 0.5352\nH 3.43 0.3431 1.338\nH 2.644 -1.542 -0.1625\nH 3.436 -0.1851 -1.027\nH 1.203 0.7551 - 1.677\nH 1.508 -0.8856 -2.298\nH 0.2104 -1.583 0.04887\nH -0.5783 - 1.375 -1.543\nH -1.172 0.9684 -1.034\nH -0.3301 0.8759 0.5549\nH - 1.891 -0.9536 1.281\nH -2.738 -0.8594 -0.3097\nH -3.962 0.4059 1.441\nH -2.507 1.426 1.755\nH -3.358 1.522 0.158" }</pre> <p>>>>>>>> EXECUTING FUNCTION query_DFT...</p> <p>[...]</p> <p>E1 = -997.6630956252707 E_coul = 406.0492308406011 Extra cycle E = -274.409942204028 delta_E = -5.68e-12 g = 1.45e-06 ddm = 5.38e-06 CPU time for scf_cycle 59.21 sec, wall time 3.03 sec CPU time for SCF 59.28 sec, wall time 3.05 sec converged SCF energy = -274.409942204028</p> <p>***** Response from calling function "query_DFT" *****</p> <p>-274.4099422040276</p>	Get coordinates of CCCCCC, and run DFT
<p>Coordinate retriever:</p> <p>***** Suggested function Call: coords_from_SMILES *****</p> <p>Arguments:</p> <pre>{ "SMILES": "CCCCOCC" }</pre>	Get coordinates of CCCCCC, and run DFT

Table 8. continued

Conversation	Notes
<pre> >>>>>>> EXECUTING FUNCTION coords_from_SMILES... Chatbot: **** Response from calling function "coords_from_SMILES" **** [C 3.481 0.4249 -0.2857,'C 2.112 0.8256 0.255','C 1.052 -0.2277 -0.08758','C -0.3203 0.172 0.4532','O -1.257 -0.8308 0.1181','C -2.564 -0.4335 0.4876','C -3.297 0.1551 -0.7131','H 3.445 0.3256 -1.392','H 3.805 -0.5403 0.1593','H 4.227 1.204 -0.02324','H 1.824 1.805 - 0.1858','H 2.181 0.9428 1.359','H 0.9859 -0.3436 -1.192','H 1.344 - 1.206 0.3525','H -0.2527 0.275 1.56','H -0.6083 1.151 0.005801','H - 2.594 0.285 1.341','H -3.116 -1.338 0.8194','H -3.305 -0.5765 -1.549', 'H -2.798 1.087 -1.051','H -4.344 0.3929 -0.4327'] ***** **** Suggested function Call: query_DFT **** Arguments: { "coordinates": "C 3.481 0.4249 -0.2857\nC 2.112 0.8256 0.255\nC 1.052 -0.2277 -0.08758\nC -0.3203 0.172 0.4532\nO -1.257 -0.8308 0.1181\nC -2.564 -0.4335 0.4876\nC -3.297 0.1551 -0.7131\nH 3.445 0.3256 -1.392\nH 3.805 -0.5403 0.1593\nH 4.227 1.204 -0.02324\nH 1.824 1.805 -0.1858\nH 2.181 0.9428 1.359\nH 0.9859 -0.3436 - 1.192\nH 1.344 -1.206 0.3525\nH -0.2527 0.275 1.56\nH -0.6083 1.151 0.005801\nH -2.594 0.285 1.341\nH -3.116 -1.338 0.8194\nH - 3.305 -0.5765 -1.549\nH -2.798 1.087 -1.051\nH -4.344 0.3929 - 0.4327" } ***** >>>>>>> EXECUTING FUNCTION query_DFT... [...] E1 = -1047.5752594421979 E_coul = 420.8765995972293 Extra cycle E= -310.227522287807 delta_E= -3.98e-12 g = 1.16e-06 ddm = 5.03e-06 CPU time for scf cycle 68.75 sec, wall time 3.51 sec CPU time for SCF 68.86 sec, wall time 3.54 sec converged SCF energy = -310.227522287807 **** Response from calling function "query_DFT" **** -310.22752228780735 </pre>	
<pre> Coordinate retriever: **** Suggested function Call: coords_from_SMILES **** Arguments: { "SMILES": "CCCCNCC" } >>>>>>> EXECUTING FUNCTION coords from SMILES... **** Response from calling function "coords_from_SMILES" **** [C -3.667 0.4902 -0.2323,'C -2.421 -0.338 -0.5355','C -1.151 0.3803 -0.06634','C 0.09811 -0.4524 -0.3695','N 1.299 0.269 0.06676','C 2.505 -0.4832 -0.298','C 3.751 0.3111 0.07632','H -3.758 0.6629 0.8612','H -4.569 -0.05361 -0.5827','H -3.616 1.469 -0.7556','H -2.509 -1.32 - 0.02114','H -2.368 -0.518 -1.631','H -1.213 0.5621 1.029','H -1.068 1.362 -0.5827','H 0.1453 -0.6358 -1.467','H 0.02723 -1.432 0.1556','H 1.269 0.3373 1.112','H 2.527 -1.464 0.2291','H 2.526 -0.665 -1.396', 'H 3.784 0.4793 1.174','H 3.748 1.292 -0.4442','H 4.659 -0.2529 - 0.2257'] </pre>	Get coordinates of CCCCNC, and run DFT

Table 8. continued

Conversation	Notes
Chatbot: ***** Suggested function Call: query_DFT ***** Arguments: { "coordinates": "C -3.667 0.4902 -0.2323\nC -2.421 -0.338 -0.5355\nC - 1.151 0.3803 -0.06634\nC 0.09811 -0.4524 -0.3695\nN 1.299 0.269 0.06676\nC 2.505 -0.4832 -0.298\nC 3.751 0.3111 0.07632\nH -3.758 0.6629 0.8612\nH -4.569 -0.05361 -0.5827\nH -3.616 1.469 -0.7556\nH - 2.509 -1.32 -0.02114\nH -2.368 -0.518 -1.631\nH -1.213 0.5621 1.029\nH -1.068 1.362 -0.5827\nH 0.1453 -0.6358 -1.467\nH 0.02723 - 1.432 0.1556\nH 1.269 0.3373 1.112\nH 2.527 -1.464 0.2291\nH 2.526 -0.665 -1.396\nH 3.784 0.4793 1.174\nH 3.748 1.292 - 0.4442\nH 4.659 -0.2529 -0.2257" } ***** >>>>>>> EXECUTING FUNCTION query_DFT... [..] E1 = -1008.918528295655 E_coul = 407.03364174745474 Extra cycle E= -290.399363184172 delta_E= -1.63e-11 g = 2.25e-06 ddm = 1.09e-05 CPU time for scf_cycle 63.95 sec, wall time 3.29 sec CPU time for SCF 64.02 sec, wall time 3.31 sec converged SCF energy = -290.399363184172 ***** Response from calling function "query_DFT" ***** -290.3993631841716	
Planner: The energies of the molecules with C, O and N are -274.41, -310.23 and -290.4 respectively. Thus, the molecule with O (CCCCOCC) leads to the lowest energy structure.	The Planner considers all simulation results, analyzes and answers the question.
Chatbot: The molecule CCCCCOCC has the lowest energy structure. TERMINATE	The Chatbot provides the final answer after checking the results.

"The set of agents include the "User" (human who poses the question), a "Planner" who develops a plan to answer the question, a "Coordinate retriever" that can retrieve coordinates from SMILES coordinates or its own knowledge, and a "Chatbot" with broad skills and execution power. The agents have access to a Python function that converts SMILES code into atom types and molecular coordinates (the molecular structure), as well as a Python function that conducts a DFT simulation given a molecular structure as an input. The question posed here is "Consider this molecule in SMILES representation: CCCCXCC, where X is one of (C, O, N). Which of these options leads to the lowest energy structure?" The agents devise a plan, carry out the simulations, analyze the data, and provide the final response to the user.

- "Modeling Expert"
- "Reviewer" who critically assesses the responses and integrates knowledge.

The boss is a proxy for the human user, who asks the questions and gives approval and input at different stages of the conversation; in this case, we set up the conversation so that human input is sought periodically to allow for follow-up questions and refinements of the answer. Each of the agents can use experts with deep technical knowledge for additional information, and we define three expert agents using RAG with access to specific knowledge (see [Materials and Methods](#) for details). [Table 9](#) shows the results of the agent set shown in [Figure 7](#), used to answer a complex design question: "Design molybdenene 2D materials by using ideas of flaw-tolerance in α -helical protein meshes. Develop three design principles that incorporate important mechanisms."

2.3.3. Agent-Based Force-Field Development: Coding, Execution, Data Analysis, File Saving, and Plotting. The final example provides the most powerful illustration of how LLMs can not only answer complex questions via a team of agents but also write *de novo* code that is utilized to develop an answer by executing it and collecting all relevant data via an active learning process by which physical ground truths are generated via DFT modeling on the fly. Sophisticated LLMs can also write and execute code and need not rely on predetermined functions, as in the previous example. [Figure 8](#) shows an experiment of using a set of two LLM agents, each with specific expertise and capabilities, to write, execute code, and analyze the

data produced. Here, we exemplify this approach in automatic force-field fitting. The task given is

Use PySCF to calculate the energy of an O₂ molecule for varying bond lengths between 0.7 and 1.8 Å in increments of 0.1 Å. Fit a piecewise spline curve to the data. Then, the energy over the bond length for the data and the fitted potential was plotted and saved as plot_O2_spline_fit_potential.svg. Save the parameters of the spline curve fit into a file in JSON format. Save the calculation results in a file in JSON format. What is the lowest energy configuration?

Giving a general task, the set of agents automatically defines the problem and strategy, writes code, executes code, and provides the required output: A graph showing the raw energy vs geometry data, JSON file outputs of energies, a spline fit to the data, including saving the force-field parameters as a JSON file. The final output created by the set of two agents ([Figure 8a](#)) is

Great! The code was executed successfully.

The lowest energy configuration for the O₂ molecule is at a bond length of approximately 1.3 Å with an energy of -148.08420987269093 Hartree.

The plot of the energy over the bond length for the data and the fitted potential has been saved as 'plot_O2_spline_fit_potential.svg'.

The parameters of the spline curve fit have been saved in a file named 'spline_params.json' in JSON format.

The calculation results have been saved in a file named 'calculation_results.json' in JSON format.

Please check these files in your current directory.

Table 9. Agent-Based Modeling Applied Here to Answer a Complex Design Question: “Design Molybdenene 2D Materials by Using Ideas of Flaw-Tolerance in α -Helical Protein Meshes. Develop three design principles that incorporate important mechanisms.”^a

Conversation	Notes
<p>Boss: Design molybdenene 2D materials by using ideas of flaw-tolerance in alpha-helical protein meshes. Develop three design principles that incorporate important mechanisms.</p> <p>Senior Engineer: Sure, here is the plan:</p> <ol style="list-style-type: none"> Researcher's task: The researcher will use the 'retrieve_content_protein' function to gather information about the flaw-tolerance in alpha-helical protein meshes. They will also use the 'retrieve_content_moly' function to gather information about molybdenene 2D materials. The researcher will then synthesize this information to develop three design principles that incorporate important mechanisms. Modeling Expert's task: The modeling expert will use the 'retrieve_content_book' function to gather information about atomistic and multiscale modeling, especially materials deformation and failure. They will then use this information to suggest suitable modeling methods for the design of molybdenene 2D materials based on the principles developed by the researcher. Reviewer's task: The reviewer will critically evaluate the design principles developed by the researcher and the modeling methods suggested by the modeling expert. They will provide feedback and suggest improvements if necessary. <p>After these steps, I will consolidate all the information and present it to you for approval.</p> <p>Does this plan meet your expectations, Boss?</p>	<p>Initial question is posed by the Boss, on behalf of the human user.</p> <p>The Senior Engineer develops a plan including detailed retrieval strategies for key information.</p> <p>As instructed, the Boss is asked to review and approve the plan.</p>
<p>Boss: <i>This plan is excellent. Please proceed to answer the question.</i></p>	<p>The human user approves the plan (potentially, the human user can provide alternative suggestions or critically assess, in which case the Senior Engineer would revise the plan).</p>
<p>Senior Engineer: ***** Suggested function Call: retrieve_content_protein ***** Arguments: { "message": "flaw-tolerance in alpha-helical protein meshes" } ***** >>>>>>> EXECUTING FUNCTION retrieve_content_protein... doc_ids: [['doc_13', 'doc_2']] Adding doc_id doc_13 to context. Adding doc_id doc_2 to context. Modeling_expert (to chat_manager): ***** Response from calling function "retrieve_content_protein" ***** [...retrieves chunks of text from the protein knowledge database...]</p> <p>***** Suggested function Call: retrieve_content_moly ***** Arguments: { "message": "molybdenene 2D materials" } >>>>>>> EXECUTING FUNCTION retrieve_content_moly... doc_ids: [['doc_1', 'doc_8']] Adding doc_id doc_1 to context. Adding doc_id doc_8 to context. Modeling_expert (to chat_manager): ***** Response from calling function "retrieve_content_moly" ***** [...retrieves chunks of text from the molybdene dataset...]</p>	<p>The Senior Engineer begins to query the expert agents – first, the protein expert assistant, then the molybdenene expert assistant.</p>
<p>Modeling expert: ***** Suggested function Call: retrieve_content_book ***** Arguments: { "message": "atomistic and multiscale modeling, materials deformation and failure" } >>>>>>> EXECUTING FUNCTION retrieve_content_book... doc_ids: [['doc_9', 'doc_492']] Adding doc_id doc_9 to context. Adding doc_id doc_492 to context. Reviewer (to chat_manager): [...retrieves chunks of text from the book dataset...]</p>	<p>The modeling expert receives feedback from the modeling expert assistant.</p>

Table 9. continued

Conversation	Notes
<p>Reviewer: Based on the information gathered, here are three design principles for molybdenene 2D materials using ideas of flaw-tolerance in alpha-helical protein meshes:</p> <ol style="list-style-type: none"> 1. Energy Dissipation: Alpha-helical protein meshes have a high energy dissipation ability due to the structure of alpha-helical proteins which combine a large array of small groups of H-bonds. These unfold concurrently in groups of 3–4 at relatively small force levels, providing a strongly nonlinear material behavior. This principle can be applied to molybdenene 2D materials by designing them in a way that allows for high energy dissipation, perhaps through the incorporation of similar small groups of bonds that can unfold under stress. 2. Self-Healing: The use of H-bonds in alpha-helical protein meshes renders the structure self-healing, as H-bonds can reform at moderate temperatures and restore the initial structure even after severe deformation. This principle can be applied to molybdenene 2D materials by incorporating elements that can reform after deformation, thereby increasing the material's durability and lifespan. 3. Flaw-Tolerance: Alpha-helical protein meshes do not display a strong stress concentration at corners of cracks, which is a common cause of failure in conventional solids. This is due to the ability of each filament in the mesh to independently stretch without affecting neighboring bonds. This principle can be applied to molybdenene 2D materials by designing them in a way that minimizes stress concentration at potential points of failure. <p>To model these principles, multiscale modeling techniques can be used. These techniques allow a rigorous linking of material properties from quantum mechanics to mesoscale and macroscale, which could be crucial in designing molybdenene 2D materials that incorporate the principles derived from alpha-helical protein meshes.</p> <p>In addition, the use of computational experiments can help in the design and testing of these materials. For instance, the behavior of the materials under different stress conditions can be simulated to ensure that they adhere to the principles of energy dissipation, self-healing, and flaw-tolerance.</p> <p>Finally, experimental techniques such as TEM and nanoindentation can be used to analyze the structure and mechanical properties of the designed materials. These techniques can provide valuable insights into the effectiveness of the design principles and can guide further improvements.</p>	
<p>Boss: <i>Thank you. Can you elaborate a bit more on how flaw-tolerance can be achieved in molybdenene, especially in terms of structural design?</i></p>	<p>Human user input is sought, with the possibility to add additional questions. Here, the human user asks for additional details on one of the aspects of the answer.</p>
<p>Senior Engineer: Certainly, Boss.</p> <p>Flaw-tolerance in molybdenene 2D materials can be achieved by borrowing the concept of independent stretching of filaments from alpha-helical protein meshes. In these protein meshes, each filament can stretch independently without affecting neighboring bonds. This is possible due to the absence of immediate interactions between individual filaments in the network, aside from cross-links at node points of the mesh. This design allows the network to avoid strong stress concentration at the corners of cracks, which is a common cause of failure in conventional solids.</p> <p>In the context of molybdenene 2D materials, this principle can be applied by designing the material in a way that allows for independent movement of its constituent parts under stress. This could potentially be achieved by incorporating a network-like structure into the material, similar to the mesh structure of alpha-helical protein networks.</p> <p>In this network, the constituent parts of the molybdenene material could be designed to move independently under stress, thereby dissipating energy and avoiding the concentration of stress at any particular point. This would require careful consideration of the bonds between the constituent parts of the material, to ensure that they allow for independent movement while still maintaining the overall integrity of the material.</p> <p>Furthermore, the use of atomistic and multiscale modeling can help in designing such a structure. These modeling techniques can provide detailed insights into the behavior of the material at different scales, from the atomic level to the macroscopic level. This can guide the design process and help in achieving a flaw-tolerant structure.</p> <p>I hope this provides a clearer picture of how flaw-tolerance can be achieved in molybdenene 2D materials. If you have any further questions, please let me know.</p>	<p>The Senior Engineer provides additional details, as requested by the human user.</p>
<p>Boss: <i>Thank you. Can you comment on how such a network-like structure could be manufactured in molybdenene?</i></p>	<p>Another follow-up question is posed by the human user.</p>

Table 9. continued

Conversation	Notes
<p>Modeling expert: Manufacturing a network-like structure in molybdenene 2D materials could be a challenging task due to the atomic scale of the material. However, advanced nanotechnology techniques could potentially be used to achieve this.</p> <p>One possible method could be the use of a bottom-up approach, such as chemical vapor deposition (CVD), where individual atoms or molecules are assembled to form the desired structure. In the case of molybdenene, this could involve the controlled deposition of molybdenum atoms in a way that forms a network-like structure. This would require precise control over the deposition process to ensure that the atoms are arranged in the desired pattern.</p> <p>Another possible method could be the use of a top-down approach, such as lithography or etching, where a bulk material is carved to create the desired structure. This could involve the use of advanced lithography techniques to carve a network-like structure into a bulk molybdenene material. However, this method could be more challenging due to the difficulty of carving such a fine structure into the material.</p> <p>Regardless of the method used, the manufacturing process would likely require the use of advanced characterization techniques, such as transmission electron microscopy (TEM), to verify the structure of the material. Additionally, computational modeling could play a crucial role in guiding the manufacturing process and predicting the behavior of the material.</p> <p>It's important to note that these are just potential methods and the actual feasibility would depend on various factors, including the specific properties of molybdenene and the available manufacturing technology.</p>	The Modeling Expert provides an answer.
Boss: Thank you!	The conversation ends once the human user approves.

“The set of agents include the “Boss” (human who poses the question), a “Senior Engineer” who develops a plan to answer the question, a “Modeling Expert”, and a “Reviewer” who critically assesses the responses and integrates knowledge. The boss is the human user who asks the questions and gives approval and input at different stages of the conversation. For instance, this includes approval of the plan and follow-up questions. Human input in the conversation below is marked in *italics*. Each of the agents can use experts with deep technical knowledge for additional information (see Figure 7 for a flowchart and overview of the approach). Below is a summary of how the conversation unfolds.

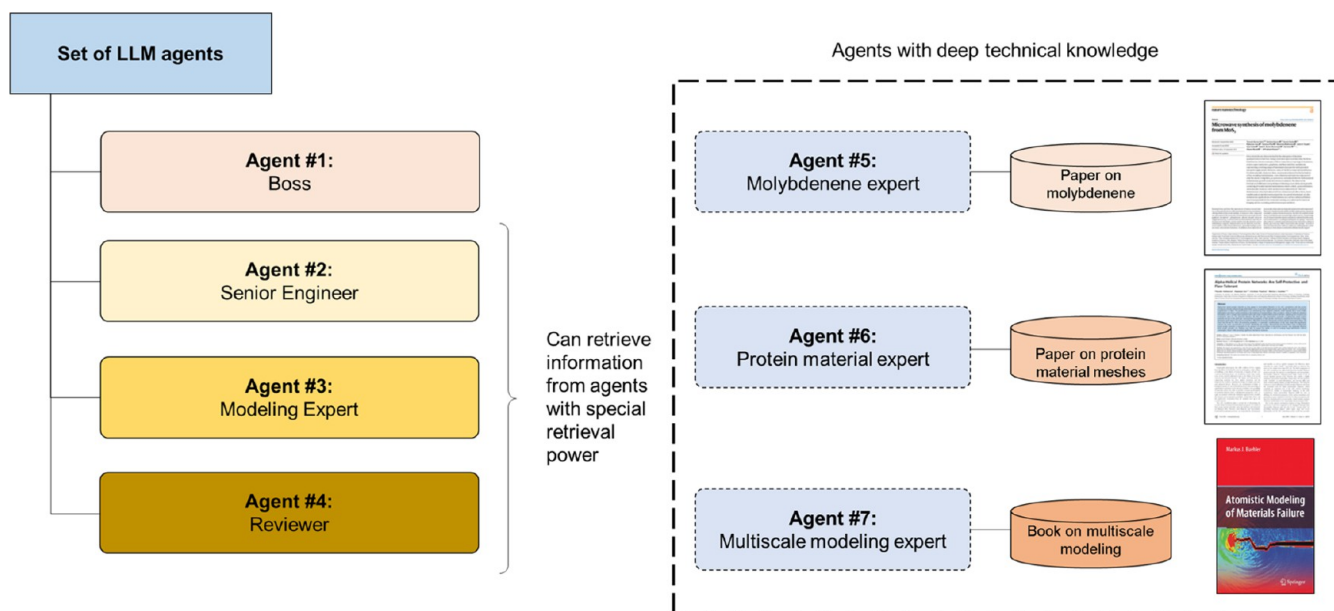
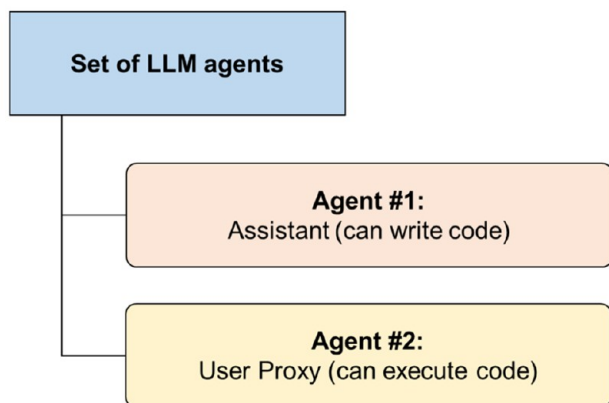


Figure 7. Using a set of seven LLM agents, each with specific expertise and capabilities, we can answer complex questions here: “Design molybdenene 2D materials by using ideas of flaw-tolerance in α -helical protein meshes. Develop three design principles that incorporate important mechanisms.”. The result of the conversation is summarized in Table 9. Note only agents #1–4 participate in the group chat conversation, whereas expert agents #5–7 are called upon by any of the agents #2–4 to retrieve information as needed from the experts. An arbitrary set of data can be made accessible to the expert agents; here, we use the molybdenene paper,⁶² the protein mesh paper,⁶⁵ and a book on multiscale modeling,⁶⁰ respectively.

A full summary of the conversation between the agents is provided in Table 10, and Figure 8b shows the key result of the experiments, including the code produced, as well as the figure generated, and the JSON data files that feature a full representation of the raw DFT data and the force-field fitting. We emphasize that no predetermined functions are used.

Instead, the LLMs write and execute all code simply driven by the task given. Such an AI system driven by an amalgamation of LLMs and other codes and methods is a modern realization of earlier concepts to develop materials design paradigms, such as the Computational Materials Design Facility (CMD⁴⁹). With the advent of increasingly sophisticated LLMs it is likely that

a.

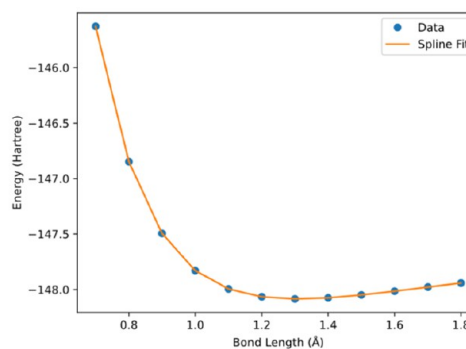


Use PySCF to calculate the energy of a O2 molecule, for varying bond lengths between 0.7 and 1.8 Angstroms in increments of 0.1 Angstroms. Fit a piecewise spline curve to the data. Then plot the energy over the bond length for the data and the fitted potential, and save as `plot_O2_spline_fit_potential.svg`. Save the parameters of the spline curve fit into a file, in JSON format. Save the calculation results in a file, in JSON format. What is the lowest energy configuration?

b.

```

1 import numpy as np
2 from scipy.interpolate import CubicSpline
3 import matplotlib.pyplot as plt
4 from pyscf import gto, dft
5 import json
6
7 bond_lengths = np.arange(0.7, 1.9, 0.1)
8 energies = []
9
10 for bond_length in bond_lengths:
11     mol = gto.M(atom='O @ 0; O @ 0 {}'.format(bond_length), basis='sto-3g')
12     mf = dft.RKS(mol)
13     mf.xc = 'pbe'
14     energy = mf.kernel()
15     energies.append(energy)
16
17 spline = CubicSpline(bond_lengths, energies)
18
19 plt.plot(bond_lengths, energies, 'o', label='Data')
20 plt.plot(bond_lengths, spline(bond_lengths), label='Spline Fit')
21 plt.xlabel('Bond Length (Å)')
22 plt.ylabel('Energy (Hartree)')
23 plt.legend()
24 plt.savefig('plot_O2_spline_fit_potential.svg')
25 plt.show()
26
27 spline_params = {'knots': spline.k.tolist(), 'coefficients': spline.c.tolist()}
28 with open('spline_params.json', 'w') as f:
29     json.dump(spline_params, f)
30
31 calculation_results = {'bond_lengths': bond_lengths.tolist(), 'energies': energies}
32 with open('calculation_results.json', 'w') as f:
33     json.dump(calculation_results, f)
34
35 min_energy = min(energies)
36 min_bond_length = bond_lengths[np.argmin(energies)]
37
38 print('The lowest energy configuration is at a bond length of {} Å with an energy of {} Hartree.'.format(min_bond_length, min_energy))
39
  
```



```

root
├── bond_lengths [ 12 items
│   ├── 0 0.7
│   ├── 1 0.7999999999999999
│   ├── 2 0.8999999999999999
│   ├── 3 0.9999999999999999
│   ├── 4 1.0999999999999999
│   ├── 5 1.1999999999999997
│   ├── 6 1.2999999999999998
│   ├── 7 1.4
│   ├── 8 1.4999999999999998
│   ├── 9 1.5999999999999996
│   ├── 10 1.6999999999999997
│   └── 11 1.7999999999999996
├── energies [ 12 items
│   ├── 0 -145.62788167137018
│   ├── 1 -146.84681171028765
│   ├── 2 -147.4949800447528
│   ├── 3 -147.83009817046496
│   ├── 4 -147.99426355165212
│   ├── 5 -148.0650523094651
│   ├── 6 -148.08420987269093
│   ├── 7 -148.07433587538745
│   ├── 8 -148.04836422105808
│   ├── 9 -148.0143775207076
│   ├── 10 -147.9776738419034
│   └── 11 -147.9415904217127
├── knots [ 12 items
│   ├── 0 0.7
│   ├── 1 0.7999999999999999
│   ├── 2 0.8999999999999999
│   ├── 3 0.9999999999999999
│   ├── 4 1.0999999999999999
│   ├── 5 1.1999999999999997
│   ├── 6 1.2999999999999998
│   ├── 7 1.4
│   ├── 8 1.4999999999999998
│   ├── 9 1.5999999999999996
│   ├── 10 1.6999999999999997
│   └── 11 1.7999999999999996
└── coefficients [ 4 items
    ├── 0 [ 11 items
    │   ├── 0 -47.44335906256796
    │   ├── 1 -47.44335906256796
    │   ├── 2 -20.494700386528397
    │   ├── 3 -12.675303619277317
    │   ├── 4 -6.38020628718909
    │   ├── 5 -3.5493000190047077
    │   ├── 6 -2.0222276946185467
    │   ├── 7 -1.2956927059586987
    │   ├── 8 -0.8776124863133751
    │   ├── 9 -0.49192491618292783
    │   └── 10 -0.49192491618291884
    ├── 1 [ 11 items
    ├── 2 [ 11 items
    └── 3 [ 11 items
  
```

JSON files produced based on the analysis/development

Figure 8. Using a set of two LLM agents, each with specific expertise and capabilities, to write, execute code, correct errors, and analyze the data produced used here to develop a computational design framework based on first-principles simulations. Panel a shows the implementation of the model (left) and the task given (right). Panel b shows three key elements of the results produced. Giving a task, the set of agents automatically defines the problem and strategy, writes code (shown on the left), executes code, and provides the required output: A graph showing the raw energy vs geometry data (top right), JSON file outputs of energies (lower right), a spline fit to the data (top right), including saving the force-field parameters as a JSON file (lower right). The plot is contained in a file named `plot_O2_spline_fit_potential.svg`, as requested in the task. The resulting code is stored in a folder (as specified when the agent is defined) so that it can be retrieved later and used outside of the AI agent chat via human users. This modality allows users to expand the solution generated by the agents toward other related use cases with additional domain knowledge.

Table 10. Agent-Based Modeling, Applied Here to Write Code, Execute, and Plot the Results⁴

Conversation	Notes
<p>User: Use PySCF to calculate the energy of a O2 molecule, for varying bond lengths between 0.7 and 1.8 Angstroms in increments of 0.1 Angstroms. Fit a piecewise spline curve to the data. Then plot the energy over the bond length for the data and the fitted potential, and save as plot_O2_spline_fit_potential.svg. Save the parameters of the spline curve fit into a file, in JSON format. Save the calculation results in a file, in JSON format. What is the lowest energy configuration?</p>	The User posts the tasks and initiates the conversation.
<p>Assistant: To calculate the energy of an O2 molecule using PySCF and fit a piecewise spline curve to the data, we can follow these steps:</p> <ol style="list-style-type: none"> 1. Install the necessary packages: <ul style="list-style-type: none"> - PySCF: `pip install pyscf` - SciPy: `pip install scipy` - NumPy: `pip install numpy` - Matplotlib: `pip install matplotlib` 2. Import the required modules: <pre> python import numpy as np from scipy.interpolate import CubicSpline import matplotlib.pyplot as plt from pyscf import gto, dft import json </pre> 3. Define the bond lengths and an empty list to store the energy values: <pre> python bond_lengths = np.arange(0.7, 1.9, 0.1) energies = [] </pre> 4. Perform the calculations for each bond length: <pre> python for bond_length in bond_lengths: mol = gto.M(atom='O 0 0 0; O 0 0 {}'.format(bond_length), basis='sto-3g') mf = dft.RKS(mol) mf.xc = 'pbe' energy = mf.kernel() energies.append(energy) </pre> 5. Fit a piecewise spline curve to the data: <pre> python spline = CubicSpline(bond_lengths, energies) </pre> 6. Plot the energy over the bond length for the data and the fitted potential: <pre> python plt.plot(bond_lengths, energies, 'o', label='Data') plt.plot(bond_lengths, spline(bond_lengths), label='Spline Fit') plt.xlabel('Bond Length (Å)') plt.ylabel('Energy (Hartree)') plt.legend() plt.savefig('plot_O2_spline_fit_potential.svg') plt.show() </pre> 7. Save the parameters of the spline curve fit into a file in JSON format: <pre> python spline_params = {'knots': spline.x.tolist(), 'coefficients': spline.c.tolist(), 'degree': spline.k} with open('spline_params.json', 'w') as f: json.dump(spline_params, f) </pre> 	The Assistant provides the first response, including a plan to install packages, import models and integrating it as a code for execution.

Table 10. continued

Conversation	Notes
<p>8. Save the calculation results in a file in JSON format:</p> <pre> '''python calculation_results = {'bond_lengths': bond_lengths.tolist(), 'energies': energies} with open('calculation_results.json', 'w') as f: json.dump(calculation_results, f) ''' </pre> <p>To find the lowest energy configuration, we can use the `min` function to find the minimum energy value and its corresponding bond length:</p> <pre> '''python min_energy = min(energies) min_bond_length = bond_lengths[np.argmin(energies)] ''' </pre> <p>Finally, print the lowest energy configuration:</p> <pre> '''python print("The lowest energy configuration is at a bond length of {} Å with an energy of {} Hartree.".format(min_bond_length, min_energy)) ''' </pre> <p>Please note that the code assumes you have the necessary packages installed and that you have the required permissions to save files in the current directory.</p>	
<pre> >>>>>>> EXECUTING CODE BLOCK 0 (inferred language is python)... >>>>>>> EXECUTING CODE BLOCK 1 (inferred language is python)... User: exitcode: 1 (execution failed) Code output: Traceback (most recent call last): File "", line 1, in <module> bond_lengths = np.arange(0.7, 1.9, 0.1) NameError: name 'np' is not defined </pre>	<p>The User agent tries to run the code; however, error messages are retrieved.</p>
<p>Assistant: It seems that the `numpy` module (`np`) is not imported. Let's modify the code to include the necessary import statement for `numpy`:</p> <pre> '''python import numpy as np from scipy.interpolate import CubicSpline import matplotlib.pyplot as plt from pysef import gto, dft import json bond_lengths = np.arange(0.7, 1.9, 0.1) energies = [] for bond_length in bond_lengths: mol = gto.M(atom='O 0 0 0; O 0 0 {}'.format(bond_length), basis='sto-3g') mf = dft.RKS(mol) mf.xc = 'pbe' energy = mf.kernel() energies.append(energy) spline = CubicSpline(bond_lengths, energies) plt.plot(bond_lengths, energies, 'o', label='Data') plt.plot(bond_lengths, spline(bond_lengths), label='Spline Fit') plt.xlabel('Bond Length (Å)') plt.ylabel('Energy (Hartree)') plt.legend() plt.savefig('plot_O2_spline_fit_potential.svg') plt.show() spline_params = {'knots': spline.x.tolist(), 'coefficients': spline.c.tolist(), 'degree': spline.k} with open('spline_params.json', 'w') as f: json.dump(spline_params, f) ''' </pre>	<p>The agent considers the error messages and revises the code based on the feedback.</p>

Table 10. continued

Conversation	Notes
<pre> calculation_results = {'bond_lengths': bond_lengths.tolist(), 'energies': energies} with open('calculation_results.json', 'w') as f: json.dump(calculation_results, f) min_energy = min(energies) min_bond_length = bond_lengths[np.argmin(energies)] print("The lowest energy configuration is at a bond length of {} Å with an energy of {} Hartree.".format(min_bond_length, min_energy)) ''' Please try running the modified code again. Let me know if you encounter any further issues. </pre>	
<pre> >>>>>>> EXECUTING CODE BLOCK 0 (inferred language is python)... User : exitcode: 1 (execution failed) Code output: /home/mbuehler/miniconda3/envs/PyTorch/lib/python3.8/site-packages/pyscf/dft/libxc.py:771: UserWarning: Since PySCF-2.3, B3LYP (and B3P86) are changed to the VWN-RPA variant, corresponding to the original definition by Stephens et al. (issue 1480) and the same as the B3LYP functional in Gaussian. To restore the VWN5 definition, you can put the setting "B3LYP_WITH_VWN5 = True" in pyscf_conf.py warnings.warn('Since PySCF-2.3, B3LYP (and B3P86) are changed to the VWN-RPA variant, ' Traceback (most recent call last): File "", line 27, in <module> spline_params = {'knots': spline.x.tolist(), 'coefficients': spline.c.tolist(), 'degree': spline.k} AttributeError: 'CubicSpline' object has no attribute 'k' </pre>	<p>The User executes the revised, updated code. Another error message is received.</p>
<pre> Assistant: I apologize for the confusion. The `CubicSpline` object in SciPy does not have a `k` attribute. The degree of the spline is always 3 for `CubicSpline`, so we don't need to save it. Here is the corrected code: '''python import numpy as np from scipy.interpolate import CubicSpline import matplotlib.pyplot as plt from pyscf import gto, dft import json bond_lengths = np.arange(0.7, 1.9, 0.1) energies = [] for bond_length in bond_lengths: mol = gto.M(atom='O 0 0 0; O 0 0 0 {}'.format(bond_length), basis='sto-3g') mf = dft.RKS(mol) mf.xc = 'pbe' energy = mf.kernel() energies.append(energy) spline = CubicSpline(bond_lengths, energies) plt.plot(bond_lengths, energies, 'o', label='Data') plt.plot(bond_lengths, spline(bond_lengths), label='Spline Fit') plt.xlabel('Bond Length (Å)') plt.ylabel('Energy (Hartree)') plt.legend() plt.savefig('plot_O2_spline_fit_potential.svg') plt.show() spline_params = {'knots': spline.x.tolist(), 'coefficients': spline.c.tolist()} with open('spline_params.json', 'w') as f: json.dump(spline_params, f) calculation_results = {'bond_lengths': bond_lengths.tolist(), 'energies': energies} with open('calculation_results.json', 'w') as f: json.dump(calculation_results, f) min_energy = min(energies) min_bond_length = bond_lengths[np.argmin(energies)] print("The lowest energy configuration is at a bond length of {} Å with an energy of {} Hartree.".format(min_bond_length, min_energy)) ''' </pre>	<p>The Assistant again consider the error messages and further revises the code. Explanations are provided as to possible strategies to address the issue.</p>

Table 10. continued

Conversation	Notes
Please try running the modified code again. Let me know if you encounter any further issues.	
<pre>>>>>>>>> EXECUTING CODE BLOCK 0 (inferred language is python)... User: exitcode: 0 (execution succeeded) Code output: converged SCF energy = -145.62788167137 converged SCF energy = -146.846811710288 converged SCF energy = -147.494980044753 converged SCF energy = -147.830098170465 converged SCF energy = -147.994263551652 converged SCF energy = -148.065052309465 converged SCF energy = -148.084209872691 converged SCF energy = -148.074335875387 converged SCF energy = -148.048364221058 converged SCF energy = -148.014377520708 converged SCF energy = -147.977673841903 converged SCF energy = -147.941590421713 Figure(640x480) The lowest energy configuration is at a bond length of 1.299999999999998 Å with an energy of - 148.08420987269093 Hartree.</pre>	The User executes the updated code, with successful completion.
<p>Assistant: Great! The code executed successfully.</p> <p>The lowest energy configuration for the O2 molecule is at a bond length of approximately 1.3 Å with an energy of -148.08420987269093 Hartree.</p> <p>The plot of the energy over the bond length for the data and the fitted potential has been saved as 'plot_O2_spline_fit_potential.svg'.</p> <p>The parameters of the spline curve fit have been saved in a file named 'spline_params.json' in JSON format.</p> <p>The calculation results have been saved in a file named 'calculation_results.json' in JSON format.</p> <p>Please check these files in your current directory.</p>	The Assistant uses the result of the code execution to answer additional questions (here, specifically the bond length at minimum energy, the stable configuration). It also provide a summary of all files/data produced, including the JSON files.

"The task given is "Use PySCF to calculate the energy of an O2 molecule, for varying bond lengths between 0.7 and 1.8 Å in increments of 0.1 Å. Fit a piecewise spline curve to the data. Then, plot the energy over the bond length for the data and the fitted potential and save it as plot_O2_spline_fit_potential.svg. Save the parameters of the spline curve fit into a file in JSON format. Save the calculation results in a file in JSON format. What is the lowest energy configuration?" As can be verified in the conversation, the set of agents not only write code, install any required libraries, and execute it but also correct any errors that may occur (this self-correcting behavior is key to achieve successful complex code development and execution).

such tasks can be more easily realized at scale and with full autonomy.

3. CONCLUSIONS

While fine-tuned LLMs are powerful tools, they can fall short of providing an accurate response or to provide up-to-date scientific results. Through a series of strategies, especially using Retrieval-Augmented Generation strategies (RAG), the use of Ontological Knowledge Graphs, and sets of LLMs interacting as autonomous agents that collaborate to solve tasks, we showed that we can inject knowledge into the generative process to enhance model performance and adapt it to new data and ultimately to being able to provide well-specified results. The case of Ontological Knowledge Graphs, in particular, is a powerful computational strategy to develop ontological insights akin to "ologs" as introduced in the earlier literature on category theory^{50–53} for use in materials science^{38,54,56–58,61} (see examples for such graphs pertinent to a particular scientific question in Figure 4b,c). Another avenue we discussed was the use of Tree-of-Thought sampling, which overcomes the limitation of a single, linear sampling step to obtain results and instead uses multidirectional and multidimensional sampling (Figure 5). Therein, we use the LLM itself to critique its own responses to filter out the most relevant details. This

takes advantage of the fact that LLMs are often excellent in providing first, generative results akin to brainstorming and also provide an accurate means to classify key responses. By breaking down a single task into multiple steps, we can generate better, more informed, and accurate outcomes.

We find that RAG and its variants are powerful methods that will likely play an important role in the future evolution of LLMs in engineering and science, especially when combined with nonlinear sampling methods. Other work in the use of retrieval-augmented LLM use cases, such as mathematical proof development,⁷² underscores these exciting possibilities. In particular, the benefit of using Ontological Knowledge Graphs is that they provide interpretable results that human scientists can read, understand, and use for either developing understanding, hypothesis, or solving complex tasks such as the building block replacement problem.⁵⁶ The examples shown in the paper showcase the construction of such graphs not only on the general corpus of data (Figure 3) but also specifically on the answering mechanism of a question (Figure 4b,c) that helps to elucidate insights into how questions are responded to within the context of knowledge. This not only provides insights that can be used, say, in an analysis or design process but also generates detailed insights into the mechanics by which the LLM constructs an answer. This can help understand the origin

of answers and can help to reverse engineer if a model predicts incorrect or unreliable results or if a user has questions.

Similar arguments can be made for complex nonlinear sampling strategies where the conventional linear mechanism of prompting and answer pairing is expanded to include interactive decision mechanisms, as also pointed out in ref 73, that are highly promising, especially when combined with active, interactive, or innovative data collection strategies (data collection can range from human experts to sensors to ab initio modeling). These decision mechanisms can be automated by using the LLM to judge its own responses or can be augmented by human expert input. This fundamentally changes the paradigm from a black-box retrieval system toward a more transparent tool that typically leads to more useful insights and can answer questions that cannot be reliably retrieved using single-shot answers.^{68,74,75} In addition to the Ontological Knowledge Graphs constructed (Figures 3, 4), Table 2 is a key result of this paper, showing a comparison of responses of MechGPT alone and with the use of RAG for a specific topic around flaw-tolerance in graphene. The RAG method uses the result of a Google Scholar search on the topic of “Deep learning flaw-tolerance”, downloading the most highly cited paper (by Lew et al.⁶⁴). The result showed that MechGPT alone was not able to provide correct results for this specific, recently published paper, but MechGPT with RAG was able to provide accurate results with specific details.

Another key advantage of using retrieval-augmented methods, especially when combined with Ontological Knowledge Graphs methods, is that sources used to generate responses can be identified. This information is included in the initial processing step that provides data about what document was used and, within the document, which section or paragraph was utilized. We can extract this information by examining metadata produced by the RAG algorithm or directly querying the LLM to provide this detail. The combination of a fine-tuned special purpose LLM in conjunction with RAG or KG methods is hence not only more accurate but also provides users with details on the origin and citation of concepts used. This provides a better integration with the scientific process and acknowledgment. Ontological Knowledge Graphs of ontological knowledge can also be automatically constructed to relate disparate concepts or domains of knowledge. As an example, we presented in Figure 6 the results of using MechGPT to elucidate analogies between music and proteins, identifying a rich set of insights from key connecting principles, subgraphs, and an overall visual representation of this area of scholarship. It nicely builds on and complements earlier work^{54,76–78} based on analytical or physics-based methods that attempted to generate such relationships. It also complements adversarial language-based strategies, as proposed in ref 79, with a much different focus, specifically to capture the natural language and its innate interpretability. Using the computational method proposed here, we can dive deeper into more complex cross-disciplinary connections. If the generated graph is further queried, we could also construct graphs specifically relevant to the answering mechanism of a question (akin to what we present in Figure 4b,c). While this is left to future work, it is an exciting possibility to explore cross-cutting ties in an interpretable manner.

Finally, agent-based modeling is a powerful technique that offers enhanced problem-solving capacity, as shown in various examples, including determining energetic details about molecular design and optimization (Table 8), question answering via a team of AI agents (Figure 7 and Table 9) and

de novo code generation, and execution and problem-solving (here exemplified for force-field development, see Table 10 and Figure 8). The ability to generate code, retrieve new physical data, and integrate various information for inclusion in easily readable file formats (like JSON) offers enhanced capabilities for computational materials design with an innate integration of codes that operate across modalities and scales. The ability of the system of AI agents to develop code and processes (including error correction, as shown in Table 10) strategically through collaboration (including the possibility of human expert input) offers new paradigms for material design workflows.

Much future work is necessary to explore the behavior and use cases of LLMs and how they can be made relevant for engineering and science. We stress that more direct and simpler, and perhaps more effective, use cases of transformer models exist in various specific tasks, such as property prediction and forward/inverse design tasks, as outlined in recent papers.^{15,30,59,80} Another area of great relevance is various societal implications that can also relate to the way engineers work in the future.^{73,81} So far, we deliberately focused much of the discussion and use cases of fine-tuned LLMs on open-source variants; however, other future work could explore the use of much larger LLMs such as GPT-3.5 and GPT-4, in general, or fine-tuned variants, for tasks discussed in this study. This can provide additional insights into use cases in which RAG and graph-based methods can be utilized. Careful assessment of the use of LLMs in a scientific and engineering context is critical, reflecting potentially harmful impacts that such technologies can have if used carelessly. Further investigations may also focus on the uses of larger open-source models, such as the MechGPT-70b models based on the Llama-2-70b family of models (not used here). This, together with new larger-scale fine-tunable models such as GPT-3.5-turbo or GPT-4 can open even more powerful avenues for applications of LLMs in engineering, multiscale modeling, and complex problem solving. The use of agent models is particularly useful also to infuse key physical insight; as done in the example, we use first-principles modeling using DFT to enhance the AI system with novel data. This strategy can principally be used also to conduct various other operations such as physical soundness (e.g., mass conservation), whether a result satisfies one or more principles or a partial differential equation, or constraints or design objectives. The use of natural language, combined with the capabilities of LLMs to effectively deal with numbers, data, code, and its execution, offers a wide array of possibilities in materials science and engineering and beyond, including, but not limited to, interpretability and the possibility of direct engagement with a human expert.

4. MATERIALS AND METHODS

Figure 1 provides a summary of the strategies used in this study. Figure 1a shows a general perspective of how LLMs, such as a fine-tuned model like MechGPT, use context and queries to provide an answer. In a conventional setting, LLMs are queried against their parameter-based knowledge that was developed during the initial training and/or fine-tuning stages. By developing a knowledge base from data sources, as shown in Figure 1b, we can augment the response of an LLM by providing relevant context retrieved from the knowledge base with the question to provide an answer (Figure 1c). Other mechanisms for sampling discussed in this paper include nonlinear sampling, where we repeatedly query the LLM with a context and task and iterate through multiple generations to ultimately develop an answer (see feedback loop from the answer to formulate new context and query).

4.1. Initial Fine-Tuned MechGPT Development

All work reported here is done based on fine-tuned LLMs derived from the Llama-2 architecture, an open LLM foundation model⁴⁶ developed for interactive chat modality. The original development of MechGPT is described in earlier papers.^{28,59} We provide just a brief summary here for completeness. The MechGPT model is based on a pretrained LLM that is fine-tuned with “textbook” mechanics and materials knowledge. In the first step, we use a general-purpose language model (in our case, a combination of the Llama-2 70b chat model³⁸ and GPT-3.5/4 in order to extract knowledge from the text in the form of question–answer pairs. To do this, we show the LLM sections of the source data (books, Wikipedia pages, etc.) and ask them to (1) develop a question to which the answer is the section considered and then (2) develop a concise summary of the section considered. This achieves multiple purposes, including a cleanup of the text; the question–answer pairs are then reviewed, cleaned up (remove any data with references to figures), and complemented by additional question–answer pairs curated from human feedback. The models are developed in PyTorch⁸² and implemented within the Hugging Face ecosystem. Training of the MechGPT model is performed based on the Llama-2 transformer architecture, using the OpenOrca-Platypus2–13B⁸³ as basis (license for all Llama derivative works: <https://github.com/facebookresearch/llama/blob/main/LICENSE>). The extended data set used for training the model includes ~2600 question–answer pairs from the “*Atomistic modeling of materials failure*” textbook⁶⁰ and ~5600 question–answer pairs developed from Wikipedia sources related to mechanics concepts with a total of ~8200 question–answer pairs. The maximum length of question–answer pairs in the prompt template is 768 tokens. The prompt template, using “{Question}” and “{Answer}” as stand-ins for the corresponding question–answer pair, is

User: {Question}<end_of_turn>Assistant: {Answer}<end_of_turn>

For a specific example of a training pair of question and answer:

User: How does hyperelasticity, or the elasticity of large strains, affect the dynamics of brittle fracture, and why is it important to consider this effect in understanding the behavior of materials under stress?<end_of_turn>
>Assistant: Hyperelasticity, or the elasticity of large strains, plays a crucial role in the dynamics of brittle fracture. Unlike linear elastic behavior, which assumes that materials will return to their original shape after stress, hyperelasticity takes into account the fact that materials can undergo large deformations without breaking. This means that the elastic behavior of materials under stress can be significantly different from the linear elastic behavior assumed in many existing theories of dynamic fracture. Experiments and computer simulations have shown that the crack propagation speed can be significantly reduced compared to the predictions by linear elastic theory. This discrepancy cannot be attributed solely to the presence of imperfections in real materials, as similar discrepancies are observed in molecular dynamics simulations of perfect atomic lattices. The importance of considering hyperelasticity in understanding the behavior of materials under stress lies in its ability to explain phenomena related to crack branching and dynamic crack tip instability. Hyperelasticity can also help explain why the maximum crack propagation speed observed in some experiments and simulations is significantly lower than predicted by linear elastic theory. Overall, considering hyperelasticity is essential to developing a complete understanding of the dynamics of brittle fracture. By taking into account the large deformations that materials can undergo without breaking, researchers can better understand the behavior of materials under stress and develop more accurate predictions of their behavior.<end_of_turn>

Another one:

User: How does molecular dynamics simulation differ from purely continuum mechanics theories in handling stress concentrations and describing bond breaking at the nanoscale?<end_of_turn>Assistant: Molecular dynamics simulations can handle stress concentrations and describe bond breaking at the nanoscale better than purely continuum mechanics theories. Unlike continuum mechanics, molecular dynamics simulations can accurately describe the behavior of materials at the atomic scale, where stress concentrations and bond breaking occur. This is because molecular dynamics simulations use simplistic model potentials to represent

atomic interactions, allowing for a more straightforward and accurate description of the mechanics of brittle fracture at the nanoscale. Additionally, the use of simplistic model potentials enables immediate comparison with continuum theories, which can help develop a deeper understanding of the mechanics of brittle fracture at the nanoscale. Furthermore, advances in computational power allow for modeling at length scales on the order of micrometers, making atomistic-based modeling a promising tool for future research in the area of modeling nanomechanical phenomena and linking to continuum mechanical theories.<end_of_turn>

To fine-tune the model, we used a paged 32-bit AdamW optimizer (“paged_adamw_32bit”) with a learning rate of LR = 0.0002, $\epsilon = 1E8$, with a warmup ratio of 0.03, a cosine learning rate schedule, and a gradient norm clipping of 0.3. See Figure S1 for the distribution of token length in the training set, as well as the loss over steps and learning rate (LR) over steps.

To parallelize training, we implemented the use of Hugging Face Accelerate (<https://huggingface.co/docs/accelerate/index>). The training objective used here is to maximize the likelihood of predicting the next token (that is, a letter, part of, or a word) given the previous words for the training set developed via question–answer pairs. We use Quantized Low-Rank Adaptation (QLoRA)^{84,85} to fine-tune the model by adding additional trainable layers and freezing the original pretrained model to avoid catastrophic forgetting of the original knowledge base (this approach involves freezing parameters of the original pretrained model and introducing small additional layers that consist of trainable rank decomposition matrices (added in each of the 40 transformer layers of the model), thereby significantly reducing the number of trainable parameters (LoRA rank = 64, $LoRA_{\alpha} = 16$, $LoRA_{dropout} = 0.1$, modules with adaptor include “q_proj”, “up_proj”, “o_proj”, “k_proj”, “down_proj”, “gate_proj”, and “v_proj”). This reduction leads to improved memory efficiency and a faster training throughput.

Once trained, we used the model for inference. Typical sampling temperatures are 0.4 (except for the initial thought collection in the tree search method, which uses a higher temperature to tend the output toward more creative responses).

4.2. Generative Strategies and Sampling Methods

We use the Llama Index (https://github.com/jerryliu/llama_index) as a tool to implement Retrieval-Augmented Generation, RAG, as well as the graph-based variant. RAG allows us to combine LLMs with additional, new knowledge bases, live data, or other external sources (see schematics in Figure 1b,c). In this study, we use RAG for knowledge-intensive natural language processing tasks by allowing LLMs to access data and incorporate new information. We use LangChain (<https://github.com/langchain-ai/langchain>) to implement embeddings via the “all-MiniLM-L6-v2” model (<https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2> and <https://www.sbert.net/>). The embedding model maps text sections to a 384-dimensional vector space that is then used by the Llama Index for indexing.

For the graph-based methods, we further use NebulaGraph (<https://github.com/vesoft-inc/nebula>) in conjunction with the Llama Index to conduct the experiments. NebulaGraph is an open-source distributed graph database that can handle large-scale graphs with high efficiency. In the context of the Llama Index, it provides useful features that allow the extraction of Ontological Knowledge Graphs for specific contexts and questions asked and the ability to visualize graphs using NetworkX (<https://networkx.org/>). We use GPT-3.5-turbo and GPT-4 to construct Ontological Knowledge Graphs from the raw data due to computational efficiency and the fact that a general-purpose model is suitable for this initial extraction step (principally, this can be done with any other model).

Multistep sampling inspired by the Tree-of-Thought method is implemented by successively sampling the model using the series of questions delineated in Figure 5b. The series of interactions with the model unfolds using the following prompts to answer ‘{question}’:

1. Generate a list of initial thoughts that are relevant for answering this question: ‘{question}’. Do not answer the question.
2. Read this: ‘{Response #1}’. List the most important concepts to answer the question ‘{question}’.

3. Considering ‘{Response #2}’, answer this question with a detailed response: {question}

It is noted that both nonlinear sampling and RAG can help to minimize hallucinations. Nonlinear sampling tends to be more powerful in larger models; RAG is effective, even in small models.

4.3. Agent-Based Modeling: Autonomous Interactions of Multiple LLMs to Solve Complex Multimodal Problems

Agent modeling is implemented using AutoGen as the overarching framework (<https://github.com/microsoft/autogen>).⁸⁶ This framework provides access to a Python implementation of concurrent integration of text generation, prompting, and logical control that allows us to exploit the natural text, code, and processing modality of LLMs, including the possibility to execute code. We consider a set of agents with different behaviors, properties, and capabilities (details in the text, but an overview can be seen in Figure 1d). The agents then conduct a collaborative conversation to solve a task. The conversation may or may not involve human input (if human input is prompted, the user can ask follow-up questions, clarify issues, or give further guidance to the AI system). All agents used GPT-4 as a backbone LLM.

Three example setups are used with agents that have different capabilities. In the first example (Section 2.3.1), we define four agents: “User”, “Planner”, “Coordinate retriever”, and “Chatbot” (details in the text). To facilitate the specific tasks of coordinate generation from SMILES code, we define a function “coords_from_SMILES” that is made available to the agents, with the description “With a SMILES string as input, provides atom type, and coordinates of a molecule.” We use RDKit (<https://github.com/rdkit/rdkit>) for the conversion of SMILES codes to coordinates (we first convert the SMILES codes to coordinates, then add H atoms, then relax the structure using MMFF94.⁸⁷ Further, a function is defined to conduct a DFT simulation, “query_DFT”, with the description: “With coordinates as input, calculate the energy of a molecule.” The Density Functional Theory simulations are carried out using PySCF (<https://pyscf.org/>).^{88–90} This code was chosen since it is easily executable within a Python environment (in general, any other simulation code can be used).

The Coordinate retriever is instructed, via its system message, to retrieve coordinates in a certain format:

You retrieve coordinates of molecules from SMILES strings or based on your own knowledge. The coordinates of the molecule must be provided in the following form in units of Angstrom:

```
coordinates = 'C 0.000000 0.000000 0.117790;C 0.000000 0.000000 0.117790'
```

The “Planner” has the following system message:

Planner. Suggest a plan to solve the task.

The “Chatbot” has the following system message:

You carry out energy calculations and answer the task. Reply TERMINATE expression was given when the task was done.

Finally, the agent “User” that acts on behalf of the human user is instructed as follows:

You interact with the planner to develop a plan to solve the problem.

The “TERMINATE” message is used to signal that the question has been answered; once received, the conversation ends. Tables S1–S3 provide code snippets to showcase the definition of this setup.

In the second example (Section 2.3.2), we use a total of 7 agents, as defined in Figure 7. The system message of the “Boss” is

Boss who asks questions and gives tasks. Interact with the planner to approve the plan. Reply ‘TERMINATE’ in the end when everything is done.

We further instruct the “Boss” to request human input when the “TERMINATE” message is received; that is, before the conversation is finalized, the human user has the opportunity to either end the conversation or ask a follow-up question. The “Senior Engineer” is instructed as follows:

Planner. You are a senior materials scientist with broad knowledge. Suggest a plan. Revise the plan based on feedback from the researcher, modeling expert, and reviewer and ask the boss for approval. The plan may involve a researcher who can retrieve information about materials, a modeling expert who suggests modeling methods, and a reviewer who gives critical feedback. Explain the plan first. Be clear which step is performed by

the researcher, which step is performed by the modeling expert, and which step is performed by the reviewer. Once the plan is created, ask the boss to approve.

The “Modeling expert” is instructed as follows:

Modeling expert. You followed an approved plan. You are an expert in atomistic and multiscale modeling who contributes to the discussion by providing ideas for how simulation can enhance solving the problem.

The reviewer is instructed:

Reviewer. You follow an approved plan. You are a scientific reviewer who gives critical feedback, adds new ideas, and integrates the concepts.

The expert agents defined in Figure 7 use RAG to generate responses, where each of the agents (agents #5, #6, and #7) use data generated from the raw PDF of the respective papers, converted into markup language using the Nougat⁹¹ optical character recognition (OCR) model. The system messages for the protein, molybdenene, and modeling experts are, respectively, “Assistant who has extra content retrieval power for solving difficult problems in protein materials.”, “Assistant who has extra content retrieval power for solving difficult problems in molybdenene materials and other 2D materials.” and “Assistant who has extra content retrieval power for information related to atomistic and multiscale modeling, especially related to materials failure.” We use Chroma to develop the vector index (<https://github.com/chroma-core/chroma>) with all-MiniLM-L6-v2 as an embedding model, as above.

For the results shown in Section 2.3.3, no predetermined Python functions are used. Instead, the LLMs write and execute all code simply driven by the task given (code generated is listed in the results table, including iterative error correction, and in Figure 8). To do this, we define two agents: an “Assistant” that helps develop code and answer questions and a “User” that is a proxy for the human that can autonomously execute code. Code execution is conducted directly in the local environment (optionally, code execution can also be done in a Docker container, which works in the same way but can be a bit slower since Docker images are downloaded and launched at every code execution step). As can be seen in Table 10, the agents not only write and execute code but also develop a plan to install the proper Python libraries needed for the execution. Note that we specify a folder name in which the codes generated by the agents are stored so that they can be retrieved separately and used for other purposes.

We emphasize that for the results in Tables 8, 9 and Figure 7, we use specially designed functions to (i) retrieve 3D coordinates from SMILES codes and (ii) conduct a DFT simulation based on coordinates. In the results in Figure 8 and Table 10, no such functions are necessary since we rely on the LLM’s capability to write code/input files and conduct a variety of other tasks.

■ ASSOCIATED CONTENT

Data Availability Statement

Codes, model weights, the data set, and other materials can be found at <https://github.com/lamm-mit/MeLM>.

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acseengineeringau.3c00058>.

Overview of the dataset and training of the MechGPT model (Figure S1); example code to set up a multi-agent problem solving strategy using AutoGen (Table S1); definition of the two functions and making them available to the LLMs (Table S2); definition of the four agents and their features and instructions, and starting the conversation (Table S3) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Markus J. Buehler – Laboratory for Atomistic and Molecular Mechanics (LAMM), Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States; Department

of Civil and Environmental Engineering, Department of Mechanical Engineering, and Center for Computational Science and Engineering, Schwarzman College of Computing, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States; orcid.org/0000-0002-4173-9659; Email: mbuehler@mit.edu

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acseengineeringau.3c00058>

Author Contributions

CRedit: **Markus J. Buehler** conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing-original draft, writing-review & editing.

Notes

The author declares no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by MIT's Generative AI initiative, Google, the Army Research Office (W911NF1920098 & W911NF2220213), ONR (N00014-19-1-2375 and N00014-20-1-2189), USDA (2021-69012-35978), and NIH U01. Additional support was provided by the MIT-IBM Watson AI Lab.

REFERENCES

- (1) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I., Attention is all you need, In *Advances in Neural Information Processing Systems*; Neural information processing systems foundation, 2017, pp 5999–6009. arXiv:1706.03762v5. arXiv.org e-Print archive. <https://arxiv.org/abs/1706.03762v5> (accessed June 27, 2021).
- (2) Brown, T. B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; Agarwal, S.; A., Herbert-Voss; Krueger, G.; Henighan, T.; Child, R.; Ramesh, A.; Ziegler, D. M.; Wu, J.; Winter, C.; Hesse, C.; Chen, M.; Sigler, E.; Litwin, M.; Gray, S.; Chess, B.; Clark, J.; Berner, C.; McCandlish, S.; Radford, A.; Sutskever, I.; Amodei, D. Language Models are Few-Shot Learners, *Advances in Neural Information Processing Systems*, 2020-December (2020). arXiv:2005.14165v4. arXiv.org e-Print archive. <https://arxiv.org/abs/2005.14165v4> (accessed June 26, 2023).
- (3) Chowdhery, A.; Narang, S.; Devlin, J.; Bosma, M.; Mishra, G.; Roberts, A.; Barham, P.; Chung, H. W.; Sutton, C.; Gehrmann, S.; Schuh, P.; Shi, K.; Tsvyashchenko, S.; Maynez, J.; Rao, A.; Barnes, P.; Tay, Y.; Shazeer, N.; Prabhakaran, V.; Reif, E.; Du, N.; Hutchinson, B.; Pope, R.; Bradbury, J.; Austin, J.; Isard, M.; G., Gur-Ari; Yin, P.; Duke, T.; Levskaya, A.; Ghemawat, S.; Dev, S.; Michalewski, H.; Garcia, X.; Misra, V.; Robinson, K.; Fedus, L.; Zhou, D.; Ippolito, D.; Luan, D.; Lim, H.; Zoph, B.; Spiridonov, A.; Sepassi, R.; Dohan, D.; Agrawal, S.; Omernick, M.; Dai, A. M.; Pillai, T. S.; Pellat, M.; Lewkowycz, A.; Moreira, E.; Child, R.; Polozov, O.; Lee, K.; Zhou, Z.; Wang, X.; Saeta, B.; Diaz, M.; Firat, O.; Catasta, M.; Wei, J.; K., Meier-Hellstern; Eck, D.; Dean, J.; Petrov, S.; Fiedel, N. PaLM: Scaling Language Modeling with Pathways, (2022). arXiv:2204.02311. arXiv.org e-Print archive. <http://arxiv.org/abs/2204.02311> (accessed September 23, 2023).
- (4) Taylor, R.; Kardas, M.; Cucurull, G.; Scialom, T.; Hartshorn, A.; Saravia, E.; Poulton, A.; Kerkez, V.; Stojnic, R. Galactica: A Large Language Model for Science, (2022). arXiv:2211.09085. arXiv.org e-Print archive. <http://arxiv.org/abs/2211.09085> (accessed September 23, 2023).
- (5) Ge, Y.; Hua, W.; Mei, K.; Ji, J.; Tan, J.; Xu, S.; Li, Z.; Zhang, Y. OpenAGI: When LLM Meets Domain Experts, (2023). arXiv:2304.04370. arXiv.org e-Print archive. <http://arxiv.org/abs/2304.04370> (accessed September 23, 2023).
- (6) OpenAI, GPT-4 Technical Report, (2023). arXiv:2303.08774. arXiv.org e-Print archive. <http://arxiv.org/abs/2303.08774> (accessed September 23, 2023).
- (7) Bubeck, S.; Chandrasekaran, V.; Eldan, R.; Gehrke, J.; Horvitz, E.; Kamar, E.; Lee, P.; Lee, Y. T.; Li, Y.; Lundberg, S.; Nori, H.; Palangi, H.; Ribeiro, M. T.; Zhang, Y. Sparks of Artificial General Intelligence: Early experiments with GPT-4, (2023). arXiv:2303.12712v5. arXiv.org e-Print archive. <https://arxiv.org/abs/2303.12712v5> (accessed April 23, 2023).
- (8) Nadkarni, R.; Wadden, D.; Beltagy, I.; Smith, N. A.; Hajishirzi, H.; Hope, T. Scientific Language Models for Biomedical Knowledge Base Completion: An Empirical Study, (2021). arXiv:2106.09700. arXiv.org e-Print archive. <http://arxiv.org/abs/2106.09700> (accessed September 23, 2023).
- (9) Beltagy, I.; Lo, K.; Cohan, A. SCIBERT: A pretrained language model for scientific text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*; Association for Computational Linguistics, 2019; pp 3615–3620. DOI: 10.18653/v1/d19-1371.
- (10) Schick, T.; J., Dwivedi-Yu; R., Dessi; Raileanu, R.; Lomeli, M.; Zettlemoyer, L.; Cancedda, N.; Scialom, T. Toolformer: Language Models Can Teach Themselves to Use Tools, (2023). arXiv:2302.04761. arXiv.org e-Print archive. <http://arxiv.org/abs/2302.04761> (accessed September 23, 2023).
- (11) Mialon, G.; R., Dessi; Lomeli, M.; Nalmpantis, C.; Pasunuru, R.; Raileanu, R.; B., Rozière; Schick, T.; J., Dwivedi-Yu; Celikyilmaz, A.; Grave, E.; LeCun, Y.; Scialom, T. *Augmented Language Models: a Survey*, (2023). arXiv:2302.07842. arXiv.org e-Print archive. <http://arxiv.org/abs/2302.07842> (accessed September 23, 2023).
- (12) Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Ichter, B.; Xia, F.; Chi, E.; Le, Q.; Zhou, D. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, (2022). arXiv:2201.11903. arXiv.org e-Print archive. <http://arxiv.org/abs/2201.11903> (accessed September 23, 2023).
- (13) Brodnik, N. R.; Carton, S.; Muir, C.; Ghosh, S.; Downey, D.; Echlin, M. P.; Pollock, T. M.; Daly, S. Perspective: Large Language Models in Applied Mechanics. *J. Appl. Mech.* **2023**, *90* (10), No. 101008.
- (14) Buehler, M. J. Generative pretrained autoregressive transformer graph neural network applied to the analysis and discovery of novel proteins. *J. Appl. Phys.* **2023**, *134*, No. 084902.
- (15) Luu, R. K.; Wysokowski, M.; Buehler, M. J. Generative Discovery of Novel Chemical Designs using Diffusion Modeling and Transformer Deep Neural Networks with Application to Deep Eutectic Solvents. *Appl. Phys. Lett.* **2023**, *122*, No. 234103, DOI: 10.1063/5.0155890.
- (16) Hu, Y.; Buehler, M. J. Deep language models for interpretative and predictive materials science. *APL Mach. Learn.* **2023**, *1*, No. 010901.
- (17) Azamfirei, R.; Kudchadkar, S. R.; Fackler, J. Large language models and the perils of their hallucinations. *Crit. Care* **2023**, *27*, No. 120, DOI: 10.1186/S13054-023-04393-X.
- (18) Kandpal, N.; Deng, H.; Roberts, A.; Wallace, E.; Raffel, C. Large Language Models Struggle to Learn Long-Tail Knowledge, (2022). arXiv:2211.08411. arXiv.org e-Print archive. <http://arxiv.org/abs/2211.08411> (accessed September 23, 2023).
- (19) Varshney, N.; Yao, W.; Zhang, H.; Chen, J.; Yu, D. A Stitch in Time Saves Nine: Detecting and Mitigating Hallucinations of LLMs by Validating Low-Confidence Generation, (2023). arXiv:2307.03987v2. arXiv.org e-Print archive. <https://arxiv.org/abs/2307.03987v2> (accessed September 23, 2023).
- (20) Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y. J.; Madotto, A.; Fung, P. Survey of Hallucination in Natural Language Generation. *ACM Comput. Surv.* **2023**, *55* (12), 1–38, DOI: 10.1145/3571730.
- (21) McKenna, N.; Li, T.; Cheng, L.; Hosseini, M. J.; Johnson, M.; Steedman, M. Sources of Hallucination by Large Language Models on

- Inference Tasks, (2023). arXiv:2305.14552v1. arXiv.org e-Print archive. <https://arxiv.org/abs/2305.14552v1> (accessed September 23, 2023).
- (22) Ge, Y.; Hua, W.; Mei, K.; Ji, J.; Tan, J.; Xu, S.; Li, Z.; Zhang, Y. OpenAGI: When LLM Meets Domain Experts, (2023). arXiv:2304.04370v5. arXiv.org e-Print archive. <https://arxiv.org/abs/2304.04370v5> (accessed September 23, 2023).
- (23) Harrer, S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. *eBioMedicine* **2023**, *90*, No. 104512, DOI: 10.1016/J.EBIOM.2023.104512.
- (24) Liang, Y.; Zhang, R.; Zhang, L.; Xie, P. DrugChat: Towards Enabling ChatGPT-Like Capabilities on Drug Molecule Graphs, (2023). arXiv:2309.03907v1. arXiv.org e-Print archive. <https://arxiv.org/abs/2309.03907v1> (accessed September 12, 2023).
- (25) Wolfram, S. What is ChatGPT doing... and why does it work?, Stephen Wolfram Writings. writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/ (accessed September 23, 2023).
- (26) Luu, R. K.; Buehler, M. J. Materials Informatics Tools in the Context of Bio-Inspired Material Mechanics. *J. Appl. Mech.* **2023**, *90* (9), No. 090801, DOI: 10.1115/1.4062310.
- (27) Lu, W.; Lee, N. A.; Buehler, M. J. Modeling and design of heterogeneous hierarchical bioinspired spider web structures using deep learning and additive manufacturing. *Proc. Natl. Acad. Sci.* **2023**, *120*, No. e2305273120.
- (28) Luu, R. K.; Buehler, M. J. BioinspiredLLM: Conversational Large Language Model for the Mechanics of Biological and Bio-inspired Materials, (2023). arXiv:2309.08788v1. arXiv.org e-Print archive. <https://arxiv.org/abs/2309.08788v1> (accessed September 23, 2023).
- (29) Lu, W.; Kaplan, D. L.; Buehler, M. J. Generative modeling, design and analysis of spider silk protein sequences for enhanced mechanical properties, (2023). arXiv:2309.10170v1. arXiv.org e-Print archive. <https://arxiv.org/abs/2309.10170v1> (accessed September 23, 2023).
- (30) Buehler, M. J. MeLM, a generative pretrained language modeling framework that solves forward and inverse mechanics problems. *J. Mech. Phys. Solids* **2023**, *181*, No. 105454.
- (31) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
- (32) Liu, F. Y. C.; Ni, B.; Buehler, M. J. PRESTO: Rapid protein mechanical strength prediction with an end-to-end deep learning model. *Extreme Mech. Lett.* **2022**, *55*, No. 101803.
- (33) Yu, C. H.; Chen, W.; Chiang, Y. H.; Guo, K.; Martin Moldes, Z.; Kaplan, D. L.; Buehler, M. J. End-to-End Deep Learning Model to Predict and Design Secondary Structure Content of Structural Proteins. *ACS Biomater. Sci. Eng.* **2022**, *8*, 1156–1165.
- (34) Hu, Y.; Buehler, M. J. End-to-End Protein Normal Mode Frequency Predictions Using Language and Graph Models and Application to Sonification. *ACS Nano* **2022**, *16*, 20656–20670.
- (35) Yu, C. H.; Khare, E.; Narayan, O. P.; Parker, R.; Kaplan, D. L.; Buehler, M. J. ColGen: An end-to-end deep learning model to predict thermal stability of de novo collagen sequences. *J. Mech. Behav. Biomed. Mater.* **2022**, *125*, No. 104921.
- (36) Khare, E.; Gonzalez-Obeso, C.; Kaplan, D. L.; Buehler, M. J. CollagenTransformer: End-to-End Transformer Model to Predict Thermal Stability of Collagen Triple Helices Using an NLP Approach. *ACS Biomater. Sci. Eng.* **2022**, *8*, 4301.
- (37) Buehler, E. L.; Buehler, M. J. End-to-end prediction of multimaterial stress fields and fracture patterns using cycle-consistent adversarial and transformer neural networks. *Biomed. Eng. Adv.* **2022**, *4*, No. 100038.
- (38) Buehler, M. J. FieldPerceiver: Domain agnostic transformer model to predict multiscale physical fields and nonlinear material properties through neural ologs. *Mater. Today* **2022**, *57*, 9–25.
- (39) Yang, C.; Wang, X.; Lu, Y.; Liu, H.; Le, Q. V.; Zhou, D.; Chen, X. Large Language Models as Optimizers, (2023). arXiv:2309.03409v1. arXiv.org e-Print archive. <https://arxiv.org/abs/2309.03409v1> (accessed September 12, 2023).
- (40) Fuhr, A. S.; Sumpter, B. G. Deep Generative Models for Materials Discovery and Machine Learning-Accelerated Innovation. *Front Mater.* **2022**, *9*, No. 865270.
- (41) Ni, B.; Kaplan, D. L.; Buehler, M. J. Generative design of de novo proteins based on secondary structure constraints using an attention-based diffusion model. *Chem* **2023**, *9*, 1828.
- (42) Inguva, P.; Bhute, V. J.; Cheng, T. N. H.; Walker, P. J. Introducing students to research codes: A short course on solving partial differential equations in Python. *Educ. Chem. Eng.* **2021**, *36*, 1–11.
- (43) Milano, S.; McGrane, J. A.; Leonelli, S. Large language models challenge the future of higher education. *Nat. Mach. Intell.* **2023**, *5*, 333–334.
- (44) Lim, W. M.; Gunasekara, A.; Pallant, J. L.; Pallant, J. I.; Pechenkina, E. Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *Int. J. Manage. Educ.* **2023**, *21*, No. 100790.
- (45) OpenAI, GPT-4 Technical Report, (2023). arXiv:2303.08774v3. arXiv.org e-Print archive. <https://arxiv.org/abs/2303.08774v3> (accessed September 23, 2023).
- (46) Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; Bikel, D.; Blecher, L.; Ferrer, C. C.; Chen, M.; Cucurull, G.; Esiobu, D.; Fernandes, J.; Fu, J.; Fu, W.; Fuller, B.; Gao, C.; Goswami, V.; Goyal, N.; Hartshorn, A.; Hosseini, S.; Hou, R.; Inan, H.; Kardas, M.; Kerkez, V.; Khabsa, M.; Kloumann, I.; Korenev, A.; Koura, S.; Lachaux, M.-A.; Lavril, T.; Lee, J.; Liskovich, D.; Lu, Y.; Mao, Y.; Martinet, X.; Mihaylov, T.; Mishra, P.; Molybog, I.; Nie, Y.; Poulton, A.; Reizenstein, J.; Rungta, R.; Saladi, K.; Schelten, A.; Silva, R.; Michael, E.; Ranjan, S.; Xiaqing, S.; Tan, E.; Tang, B.; Taylor, R.; Williams, A.; Kuan, J. X.; Xu, P.; Yan, Z.; Zarov, I.; Zhang, Y.; Fan, A.; Kambadur, M.; Narang, S.; Rodriguez, A.; Stojnic, R.; Edunov, S.; Scialom, T. Llama 2: Open Foundation and Fine-Tuned Chat Models, (2023). arXiv:2307.09288v2. arXiv.org e-Print archive. <https://arxiv.org/abs/2307.09288v2> (accessed August 22, 2023).
- (47) Falcon, URL <https://falconllm.tii.ae/falcon.html> (accessed June 26, 2023).
- (48) Penedo, G.; Malartic, Q.; Hesslow, D.; Cojocar, R.; Cappelli, A.; Alobeidli, H.; Pannier, B.; Almazrouei, E.; Launay, J. The RefinedWeb Dataset for Falcon LLM: Outperforming Curated Corpora with Web Data, and Web Data Only, (2023). arXiv:2306.01116v1. arXiv.org e-Print archive. <https://arxiv.org/abs/2306.01116v1> (accessed June 26, 2023).
- (49) Buehler, M. J.; Dodson, J.; Van Duin, A. C. T.; Meulbroek, P.; Goddard, W. A. The Computational Materials Design Facility (CMDf): A powerful framework for multi-paradigm multi-scale simulations. *MRS Proc.* **2005**, *894*, No. 08940303.
- (50) Eilenberg, S.; MacLane, S. General theory of natural equivalences. *Trans. Am. Math. Soc.* **1945**, *58*, 231.
- (51) Eilenberg, S.; MacLane, S. Group Extensions and Homology. *Ann. Math.* **1942**, *43*, 757–831.
- (52) Marquis, J.-P. Category Theory, Stanford Encyclopedia of Philosophy. (2019). <https://plato.stanford.edu/entries/category-theory/> (accessed September 23, 2023).
- (53) Baez, J. C.; Stay, M. Physics, Topology, Logic and Computation: A Rosetta Stone. In *Lecture Notes in Physics*; Springer: Berlin Heidelberg, 2011; Vol. 813, pp 95–172. DOI: 10.1007/978-3-642-12821-9_2 (accessed September 23, 2023).
- (54) Giesa, T.; Spivak, D. I.; Buehler, M. J. Reoccurring Patterns in Hierarchical Protein Materials and Music: The Power of Analogies. *BioNanoScience* **2011**, *1*, 153–161, DOI: 10.1007/s12668-011-0022-5.

- (55) Wong, J. Y.; McDonald, J.; Taylor-Pinney, M.; Spivak, D. I.; Kaplan, D. L.; Buehler, M. J. Materials By Design: Merging Proteins and Music. *Nano Today* **2012**, *7* (6), 488–495.
- (56) Giesa, T.; Spivak, D. I.; Buehler, M. J. Category theory based solution for the building block replacement problem in materials design. *Adv. Eng. Mater.* **2012**, *14* (9), 810–817, DOI: 10.1002/adem.201200109.
- (57) Giesa, T.; Jagadeesan, R.; Spivak, D. I.; Buehler, M. J. Matriarch: A Python Library for Materials Architecture. *ACS Biomater. Sci. Eng.* **2015**, *1* (10), 1009–1015, DOI: 10.1021/acsbomaterials.5b00251.
- (58) Brommer, D. B.; Giesa, T.; Spivak, D. I.; Buehler, M. J. Categorical prototyping: Incorporating molecular mechanisms into 3D printing. *Nanotechnology* **2015**, *27* (2), No. 024002, DOI: 10.1088/0957-4484/27/2/024002.
- (59) Buehler, M. J. MechGPT, a language-based strategy for mechanics and materials modeling that connects knowledge across scales, disciplines and modalities. *Appl. Mech. Rev.* **2023**, 1–82.
- (60) Buehler, M. J. *Atomistic Modeling of Materials Failure*; Springer, 2008 DOI: 10.1007/978-0-387-76426-9.
- (61) Spivak, D. I.; Giesa, T.; Wood, E.; Buehler, M. J. Category theoretic analysis of hierarchical protein materials and social networks. *PLoS One* **2011**, *6* (9), No. e23911, DOI: 10.1371/journal.pone.0023911.
- (62) Sahu, T. K.; Kumar, N.; Chahal, S.; Jana, R.; Paul, S.; Mukherjee, M.; Tavabi, A. H.; Datta, A.; Dunin-Borkowski, R. E.; Valov, I.; Nayak, A.; Kumar, P. Microwave synthesis of molybdenene from MoS₂. *Nat. Nanotechnol.* **2023**, *18*, 1430–1438.
- (63) Ling, S.; Kaplan, D. L.; Buehler, M. J. Nanofibrils in nature and materials engineering. *Nat. Rev. Mater.* **2018**, *3*, No. 18016, DOI: 10.1038/natrevmats.2018.16.
- (64) Lew, A. J.; Yu, C.; Hsu, Y.; Buehler, M. J. Deep learning model to predict fracture mechanisms of graphene. *npj 2D Mater. Appl.* **2021**, *5*, No. 48, DOI: 10.1038/s41699-021-00228-x.
- (65) Ackbarow, T.; Sen, D.; Thaulow, C.; Buehler, M. J. Alpha-helical protein networks are self-protective and flaw-tolerant. *PLoS One* **2009**, *4* (6), No. e6015, DOI: 10.1371/journal.pone.0006015.
- (66) Cranford, S. W.; Buehler, M. J. *Biomateriomics*; Springer: Netherlands, 2012.
- (67) Cranford, S. W.; Buehler, M. J. *The Future of Biomateriomics. In Springer Series in Materials Science*; Springer: Netherlands, 2012; pp 425–430 DOI: 10.1007/978-94-007-1611-7.
- (68) Yao, S.; Yu, D.; Deepmind, G.; Zhao, J.; Griffiths, T. L.; Cao, Y.; Narasimhan, K. Tree of Thoughts: Deliberate Problem Solving with Large Language Models, (2023). arxiv:2305.10601v1. arXiv.org e-Print archive. <https://arxiv.org/abs/2305.10601v1> (accessed September 23, 2023).
- (69) Dhuliawala, S.; Ai, M.; E., Zürich; Komeili, M.; Xu, J.; Raileanu, R.; Li, X.; Celikyilmaz, A.; Weston, J. Chain-of-Verification Reduces Hallucination in Large Language Models, (2023). arXiv:2309.11495v1. arXiv.org e-Print archive. <https://arxiv.org/abs/2309.11495v1> (accessed September 24, 2023).
- (70) Weininger, D. SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.
- (71) Parr, R. G.; Weitao, Y. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press, 1995 DOI: 10.1093/oso/9780195092769.001.0001.
- (72) Yang, K.; Swope, A. M.; Gu, A.; Chalamala, R.; Song, P.; Yu, S.; Godil, S.; Prenger, R.; Anandkumar, A.; U., Santa Barbara; Austin, U. LeanDojo: Theorem Proving with Retrieval-Augmented Language Models, (2023). arXiv:2306.15626v1. arXiv.org e-Print archive. <https://arxiv.org/abs/2306.15626v1> (accessed September 23, 2023).
- (73) van der Zant, T.; Kouw, M.; Schomaker, L. Generative artificial intelligence. In *Studies in Applied Philosophy, Epistemology and Rational Ethics*; Springer: Berlin Heidelberg, 2013; Vol. 5, pp 107–120. DOI: 10.1007/978-3-642-31674-6_8/COVER.
- (74) Long, J. Large Language Model Guided Tree-of-Thought, (2023). arxiv:2305.08291v1. arXiv.org e-Print archive. <https://arxiv.org/abs/2305.08291v1> (accessed September 23, 2023).
- (75) Chen, W.; Ma, X.; Wang, X.; Cohen, W. W. Program of Thoughts Prompting: Disentangling Computation from Reasoning for Numerical Reasoning Tasks, (2022). arxiv:2211.12588v3. arXiv.org e-Print archive. <https://arxiv.org/abs/2211.12588v3> (accessed September 23, 2023).
- (76) Franjou, S. L.; Milazzo, M.; Yu, C.-H.; Buehler, M. J. A perspective on musical representations of folded protein nanostructures. *Nano Futures* **2021**, *5*, 012501.
- (77) Yu, C. H.; Qin, Z.; Martin-Martinez, F.; Buehler, M. J. A self-consistent sonification method to translate amino acid sequences into musical compositions and application in protein design using Artificial Intelligence. *ACS Nano* **2019**, *13*, 7471–7482.
- (78) Milazzo, M.; Anderson, G. I.; Buehler, M. J. Bioinspired translation of classical music into de novo protein structures using deep learning and molecular modeling. *Bioinspir. Biomim.* **2022**, *17*, No. 015001.
- (79) Buehler, M. J. Unsupervised cross-domain translation via deep learning and adversarial attention neural networks and application to music-inspired protein designs. *Patterns* **2023**, *4*, No. 100692.
- (80) Lew, A. J.; Buehler, M. J. Single-shot forward and inverse hierarchical architected materials design for nonlinear mechanical properties using an Attention-Diffusion model. *Mater. Today* **2023**, *64*, 10.
- (81) Baldassarre, M. T.; Caivano, D.; Fernandez Nieto, B.; Gigante, D.; Ragone, A. The Social Impact of Generative AI: An Analysis on ChatGPT. In *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*; ACM, 2023; pp 363–373 DOI: 10.1145/3582515.3609555.
- (82) Paszke, A.; Gross, S.; Bradbury, J.; Lin, Z.; Devito, Z.; Massa, F.; Steiner, B.; Killeen, T.; Yang, E. PyTorch: An Imperative Style, High-Performance Deep Learning Library, 2019, arXiv:1912.01703. arXiv.org e-Print archive. <https://arxiv.org/abs/1912.01703>.
- (83) Open-Orca/OpenOrca-Platypus2–13B Hugging Face, URL: <https://huggingface.co/Open-Orca/OpenOrca-Platypus2-13B> (accessed August 27, 2023).
- (84) Hu, E. J.; Shen, Y.; Wallis, P.; Z., Allen-Zhu; Li, Y.; Wang, S.; Wang, L.; Chen, W. LoRA: Low-Rank Adaptation of Large Language Models, (2021). arxiv:2106.09685v2. arXiv.org e-Print archive. <https://arxiv.org/abs/2106.09685v2> (accessed June 26, 2023).
- (85) Dettmers, T.; Pagnoni, A.; Holtzman, A.; Zettlemoyer, L. QLoRA: Efficient Finetuning of Quantized LLMs, (2023). arXiv:2305.14314v1. arXiv.org e-Print archive. <https://arxiv.org/abs/2305.14314v1> (accessed September 12, 2023).
- (86) Wu, Q.; Bansal, G.; Zhang, J.; Wu, Y.; Li, B.; Zhu, E.; Jiang, L.; Zhang, X.; Zhang, S.; Liu, J.; Awadallah, A. H.; White, R. W.; Burger, D.; Wang, C. AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation, (2023). arXiv:2308.08155v2. arXiv.org e-Print archive. <https://arxiv.org/abs/2308.08155v2> (accessed October 28, 2023).
- (87) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (88) Sun, Q.; Zhang, X.; Banerjee, S.; Bao, P.; Barbry, M.; Blunt, N. S.; Bogdanov, N. A.; Booth, G. H.; Chen, J.; Cui, Z. H.; Eriksen, J. J.; Gao, Y.; Guo, S.; Hermann, J.; Hermes, M. R.; Koh, K.; Koval, P.; Lehtola, S.; Li, Z.; Liu, J.; Mardirossian, N.; McClain, J. D.; Motta, M.; Mussard, B.; Pham, H. Q.; Pulkin, A.; Purwanto, W.; Robinson, P. J.; Ronca, E.; Sayfutyarova, E. R.; Scheurer, M.; Schurkus, H. F.; Smith, J. E. T.; Sun, C.; Sun, S. N.; Upadhyay, S.; Wagner, L. K.; Wang, X.; White, A.; Whitfield, J. D.; Williamson, M. J.; Wouters, S.; Yang, J.; Yu, J. M.; Zhu, T.; Berkelbach, T. C.; Sharma, S.; Sokolov, A. Y.; Chan, G. K. L. Recent developments in the PySCF program package. *J. Chem. Phys.* **2020**, *153*, No. 024109, DOI: 10.1063/5.0006074.
- (89) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K. L. PySCF: the Python-based simulations of chemistry framework. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8* (1), No. e1340, DOI: 10.1002/WCMS.1340.

- (90) Sun, Q. Libcint: An efficient general integral library for Gaussian basis functions. *J. Comput. Chem.* **2015**, *36*, 1664–1671.
- (91) Blecher, L.; Cucurull, G.; Scialom, T.; Stojnic, R.; Ai, M. Nougat: Neural Optical Understanding for Academic Documents, (2023). arXiv:2308.13418v1. arXiv.org e-Print archive. <https://arxiv.org/abs/2308.13418v1> (accessed October 11, 2023).
- (92) Buehler, M. J.; Gao, H. Dynamical fracture instabilities due to local hyperelasticity at crack tips. *Nature* **2006**, *439*, 307–310, DOI: [10.1038/nature04408](https://doi.org/10.1038/nature04408).
- (93) Buehler, M. J.; Gao, H. A mother-daughter mechanism of mode I cracks: Supersonic crack motion along interfaces of dissimilar materials. *MRS Proc.* **2005**, *904*, No. 09040305.
- (94) Buehler, M. J.; Kong, Y.; Gao, H.; Huang, Y. Self-folding and unfolding of carbon nanotubes. *J. Eng. Mater. Technol.* **2006**, *128* (1), 3–10, DOI: [10.1115/1.1857938](https://doi.org/10.1115/1.1857938).
- (95) Buehler, M. J.; Abraham, F. F.; Gao, H. Hyperelasticity governs dynamic fracture at a critical length scale. *Nature* **2003**, *426*, 141–146.
- (96) Buehler, M. J.; Hartmaier, A.; Gao, H. Hierarchical multi-scale modelling of plasticity of submicron thin metal films. *Modell. Simul. Mater. Sci. Eng.* **2004**, *12*, S391 DOI: [10.1088/0965-0393/12/4/s07](https://doi.org/10.1088/0965-0393/12/4/s07).
- (97) Buehler, M. J.; Hartmaier, A.; Gao, H. Atomistic and continuum studies of crack-like diffusion wedges and associated dislocation mechanisms in thin films on substrates. *J. Mech. Phys. Solids* **2003**, *51* (11–12), 2105–2125, DOI: [10.1016/j.jmps.2003.09.024](https://doi.org/10.1016/j.jmps.2003.09.024).
- (98) De La Rubia, T. D.; Yip, S. Scientific modeling and simulations: Advocacy of computational science (editors' preface). In *Lecture Notes in Computational Science and Engineering*; Springer: Netherlands, 2008; pp 1–2 DOI: [10.1007/978-1-4020-9741-6_1](https://doi.org/10.1007/978-1-4020-9741-6_1).
- (99) Shima, H. Buckling of Carbon Nanotubes: A State of the Art Review. *Materials* **2012**, *5*, 47.
- (100) Anderson, T. L. *Fracture Mechanics: Fundamentals and Applications*; Taylor & Francis, 2005.
- (101) Anderson, P. M.; Hirth, J. P.; Lothe, J. *Theory of Dislocations*, Third ed.; Cambridge University Press, 2017; p 1543.
- (102) Scheraga, H. A.; Khalili, M.; Liwo, A. Protein-folding dynamics: overview of molecular simulation techniques. *Annu. Rev. Phys. Chem.* **2007**, *58*, 57–83.
- (103) Gao, H.; Yao, H. Shape insensitive optimal adhesion of nanoscale fibrillar structures. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7851–7856.
- (104) Kelly, C.; Gage, M. J. Protein Unfolding: Denaturant vs. Force. *Biomedicine* **2021**, *9* (10), 1395 DOI: [10.3390/BIOMEDICINES9101395](https://doi.org/10.3390/BIOMEDICINES9101395).
- (105) Zhang, T.; Li, X.; Kadkhodaei, S.; Gao, H. Flaw insensitive fracture in nanocrystalline graphene. *Nano Lett.* **2012**, *12*, 4605–4610.
- (106) Falconer, K. J. The Hausdorff dimension of self-affine fractals. *Math. Proc. Cambridge Philos. Soc.* **1988**, *103*, 339–350.
- (107) Khalid, S.; Nazir, R. Hybrid Metal-Polymer Nanocomposites: Synthesis, Characterization, and Applications. In *Handbook of Polymer and Ceramic Nanotechnology*; Springer International Publishing, 2021; pp 1–36 DOI: [10.1007/978-3-030-10614-0_78-1](https://doi.org/10.1007/978-3-030-10614-0_78-1).
- (108) Qin, Z.; Cranford, S.; Ackbarow, T.; Buehler, M. J. Robustness-strength performance of hierarchical alpha-helical protein filaments. *Int. J. Appl. Mech.* **2009**, *01* (01), 85–112, DOI: [10.1142/S1758825109000058](https://doi.org/10.1142/S1758825109000058).