PNAS RESEARCH ARTICLE | ENGINEERING

Automating alloy design and discovery with physics-aware multimodal multiagent AI

Alireza Ghafarollahi[a] and Markus J. Buehler[a,b,1]

Edited by Yonggang Huang, Northwestern University–Evanston, Glencoe, IL; received July 13, 2024; accepted November 13, 2024

The design of new alloys is a multiscale problem that requires a holistic approach that involves retrieving relevant knowledge, applying advanced computational methods, conducting experimental validations, and analyzing the results, a process that is typically slow and reserved for human experts. Machine learning can help accelerate this process, for instance, through the use of deep surrogate models that connect structural and chemical features to material properties, or vice versa. However, existing data-driven models often target specific material objectives, offering limited flexibility to integrate out-of-domain knowledge and cannot adapt to new, unforeseen challenges. Here, we overcome these limitations by leveraging the distinct capabilities of multiple AI agents that collaborate autonomously within a dynamic environment to solve complex materials design tasks. The proposed physics-aware generative AI platform, AtomAgents, synergizes the intelligence of large language models (LLMs) and the dynamic collaboration among AI agents with expertise in various domains, including knowledge retrieval, multimodal data integration, physics-based simulations, and comprehensive results analysis across modalities. The concerted effort of the multiagent system allows for addressing complex materials design problems, as demonstrated by examples that include autonomously designing metallic alloys with enhanced properties compared to their pure counterparts. Our results enable accurate prediction of key characteristics across alloys and highlight the crucial role of solid solution alloying to steer the development of advanced metallic alloys. Our framework enhances the efficiency of complex multiobjective design tasks and opens avenues in fields such as biomedical materials engineering, renewable energy, and environmental sustainability.

materials design | multiagent system | LLM | atomistic simulations | mechanics

Significance

We construct a physics-aware AI model that integrates the advanced reasoning, rational thinking, and strategic planning capabilities of large language models with the ability to write and execute code, perform atomistic simulations to solicit new physics data, and conduct visual analysis of graphed data and molecular mechanisms. By employing a multiagent strategy, these capabilities are combined into an intelligent system designed to solve complex analysis and design tasks, as applied here to alloy design and discovery.

The continuous demand for new materials is driven by the need to address emerging technological challenges, enhance efficiencies, reduce costs, and minimize environmental impacts across a range of industries (1–4). Innovations in materials science can catalyze breakthroughs in sectors such as electronics (5), aerospace (6, 7), energy storage (8, 9), and biomedicine (10). For example, the development of lighter, stronger materials could lead to more fuel-efficient vehicles and aircraft, while advancements in semiconductor technology could revolutionize electronics through enhanced functionality and reduced energy consumption (11).

Metal alloys are indispensable for many structural applications mainly due to the critical role of defects in their crystalline lattices, such as dislocations, interfaces, crack tips, grain boundaries, precipitates, and vacancies. These defects and their interactions determine key properties like plastic flow behavior, creep, fatigue, and fracture toughness, directly impacting the material's performance (12–16). Understanding these defects and their chemical dependencies is crucial for optimizing existing alloys and designing new, high-performance materials. Experimental and computational methods are essential in this pursuit, providing detailed insights into the complex behaviors of metallic systems under varying conditions. Advanced experimental techniques reveal the formation and dynamics of defects in real-time, enhancing our understanding of material behavior under operational conditions. Atomistic simulations, including molecular dynamics (MD) (17–20) and density functional theory (DFT) (21), offer detailed models of defect behaviors at the atomic level, enabling the prediction of material behaviors under various conditions. These are complemented by physics-based theoretical models, which bridge the gap between atomic-scale phenomena and macroscopic material behavior, thereby further accelerating the material design process and providing mechanistic insights. (22–30) Recently, machine learning (ML) and AI techniques have been integrated into computational materials science, enabling property prediction and accelerating materials analysis and design (31–37). In early stages of such ML driven AI methods as applied to physics, prime applications have been as surrogate models or tools to serve specific

Author affiliations: [a]Laboratory for Atomistic and Molecular Mechanics, Massachusetts Institute of Technology, Cambridge, MA 02139; and [b]Center for Computational Science and Engineering, Schwarzman College of Computing, Massachusetts Institute of Technology, Cambridge, MA 02139

Author contributions: A.G. and M.J.B. designed research; A.G. performed research; A.G. and M.J.B. contributed new reagents/analytic tools; A.G. and M.J.B. analyzed data; and A.G. and M.J.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: mbuehler@mit.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2414074122/-/DCSupplemental.

Published January 24, 2025.

inverse problems, like the design of new microstructures to meet certain material behaviors (38–44).

Materials design is inherently a multiscale challenge, necessitating the integration of materials characteristics across different scales-from atomic interactions to macroscopic behaviors (45–47). The vast array of data generated from these diverse scales-including deformation mechanisms, mechanical and thermal properties, processing–structure–property relationships, microstructural characteristics, and chemical compositions-exists in various formats such as text, images, and tabular data. Valuable insights are also embedded in resources like books, materials databases, patents, and technical reports (13, 48–50). Moreover, designing materials with enhanced performance involves satisfying multiple criteria, a task complicated by the limitations of current modeling approaches. Physics-based models are precise but generally target specific properties and depend on computationally expensive input parameters derived from intensive simulations like DFT, and their integration requires reasoning over results, simulation strategies, and understanding of relative weaknesses and strengths of various tools. Machine learning and deep learning models complement these efforts in various ways: For instance, as surrogate models that bridge different scales enabling the exploration of massive design spaces (51), or by developing state-of-the-art ML interatomic potentials to achieve DFT accuracy with the speed of empirical potentials. (52–56).

However, a persistent frontier in materials science is the development of systems that cultivate comprehensive intelligence by automating complex materials modeling and design tasks, while leveraging a diverse array of knowledge, tools, and capabilities across different scales. These systems are crucial not only for generating novel insights into materials but also for dynamically integrating existing knowledge and developing new data. Importantly, these systems should ideally be able to iteratively refine their strategies, merging insights across disciplines to progressively evolve toward optimal solutions, along with an understanding, awareness, and ability to probe specific details of the physics. Such a broad and integrative capability enables them to continuously enhance their approaches, adapting to new data (e.g. observations) and findings, thus improving the efficiency and effectiveness of material design and discovery processes. In this study, we propose a method to achieve such an integration of external knowledge, tools, logic, reasoning, and show how a physics-aware AI can be effectively implemented through the deployment of a multimodal multiagent AI system driven by large language models (LLMs).

LLMs (57, 58) have demonstrated significant potential in various scientific and engineering domains (59, 60) such as materials (61, 62), chemistry (63), mechanics (40, 64), and proteins discovery (65). Such models, built upon attention mechanism and transformer architectures (66) have shown proficiency in complex reasoning, strategic planning, coding, and workflow development, showing promising capabilities in materials analysis and prediction applications including hypothesis generation (67), critical knowledge retrieval via in-context learning, and multimodal reasoning (68). However, they face challenges in materials design due to limitations such as the inability to perform physics-based simulations, restricted access to external sources, and reliance on potentially outdated knowledge, which may not align with the rapidly advancing field of materials science.

To extend the utility of LLMs, we propose that multiagent systems can serve as frameworks that transcend the traditional conversational functions of LLMs, with incipient applications across different domains (69–73). These systems consist of three primary components (74): The brain, perception, and action, as illustrated in Fig. 1. The brain is the core of an AI agent model and is primarily composed of a frontier LLM, undertaking essential tasks of decision-making, reasoning, and planning. The perception module gathers and processes multimodal data, while the action module implements decisions based on the brain's guidance. The proposed workflow of our multimodal multiagent system is as follows: First, the perception module perceives changes in the external environment and then converts multimodal information into an understandable representation for the agents. Next, the brain module engages in processing activities such as planning, thinking, and decision-making. Finally, the action module carries out the execution with the assistance of tools. By repeating the above process, the system of agents can collaborate, continuously get feedback, and interact with the environment.

The unique characteristics of multiagent systems make them particularly suited to address the complexities of materials design, where traditional human-centric approaches may fail. These systems can integrate diverse data modalities and extract new knowledge from a broad range of external sources, including academic literature, online databases, and cutting-edge physics simulations, as depicted in Fig. 1. Enhanced by multimodal

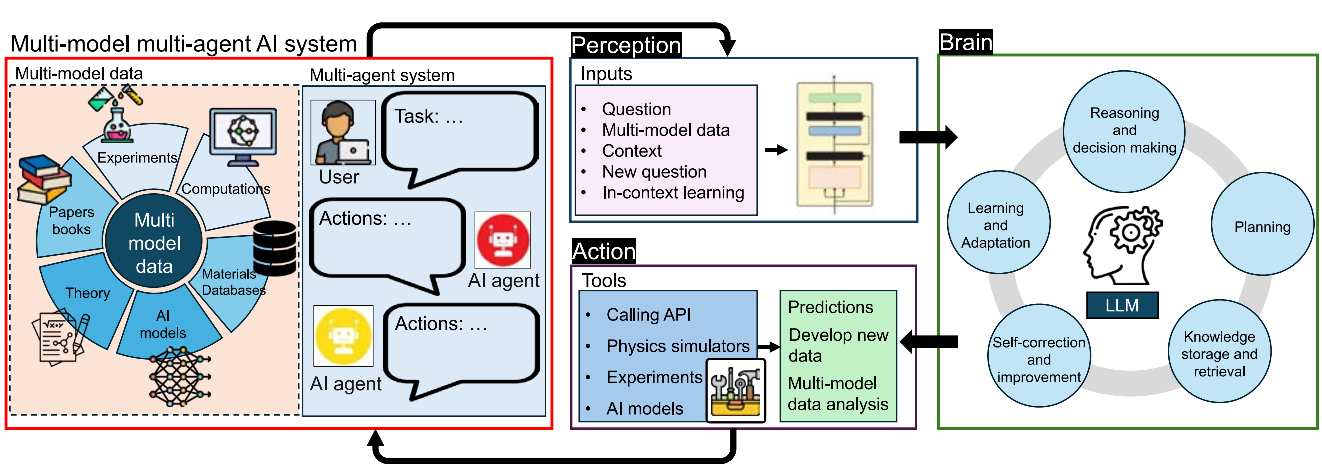


Fig. 1. Multimodal multiagent approach as a flexible modeling strategy for materials discovery, modeling, and prediction. Multiagent modeling can extend the power of large-language models by enabling the integration of multimodal data from diverse sources, including simulations, experiments, materials databases, and theoretical models.

LLMs, multiagent systems are also capable of reasoning over images, facilitating the analysis of numerous experimental and computational results in visual formats. Such capabilities ensure the continuous evolution and improvement of material design and analysis, keeping pace with new scientific discoveries and methodological advancements.

In this paper, we present a multiagent AI approach designed to address the unique challenges associated with alloy modeling and design. We propose "AtomAgents," a physics-aware multiagent framework tailored to resolve complex issues in materials design that require detailed atomistic simulations. This framework utilizes a coordinated network of multiagent systems to significantly enhance the efficiency and effectiveness of simulation processes in the development and analysis of crystalline materials at the atomic level. The primary contributions of our work are as follows:

- Integrating physics with generative AI: A deep capability to synergistically combine LLMs with detailed physics-based simulations, here demonstrated for the design of crystalline materials invoking the general-purpose open-source Large-scale Atomic-Molecular Parallel Simulator (LAMMPS) MD code (75).
- Multimodal Data Integration: Our model has the capability to integrate multimodal data from various sources, enhancing its utility and adaptability in diverse research contexts.
- Advanced Simulation Capabilities: Our model demonstrates exceptional performance in retrieving new physics through atomistic simulations, validated by several complex computational experiments.
- Reduction in Human Intervention: AtomAgents significantly reduces the need for human intervention, displaying its capacity to autonomously design complex workflows, particularly useful for high-throughput simulations.
- Accessibility for Nonexperts: Operating based on textual input, our model empowers nonexpert researchers to effectively address challenges in the realm of crystalline materials design, making advanced simulations more accessible.
- Interpretability: The interactions between agents and tools are fully traceable for interpretation and analysis of intermediate results. This allows human researchers to understand potential issues and intervene, or redirect the process if necessary.

The structure of this paper is organized as follows. In Section 1, we provide a detailed overview of the multiagent system developed in our study. Subsequently, we present a series of computational experiments designed to demonstrate the efficacy of multiagent collaboration in tackling complex tasks in alloy design and analysis involving atomistic simulations. Finally, a comprehensive discussion of the limitations and future perspectives of our approach is offered in Section 2, featuring also a critical discussion of challenges and future research opportunities.

1. Results and Discussion

The outline of our proposed multiagent model is shown in Fig. 2, illustrating the collaborative efforts of a team of AI agents to solve complex multiobjective problems in the context of alloy design and analysis requiring atomistic simulations. These

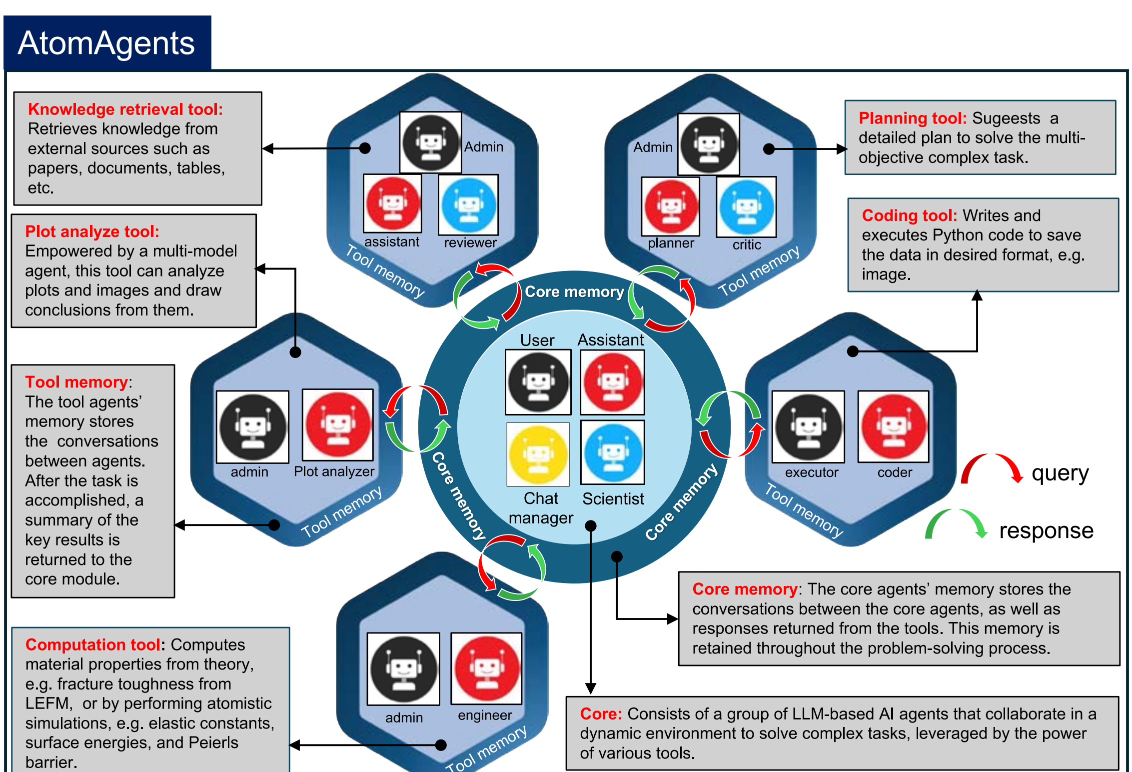


Fig. 2. AtomAgents, a physics-based generative multiagent model for automating alloy discovery and analysis with atomistic simulations. The structure of AtomAgents comprises a team of agents constructing the core who collaborate to solve complex alloy design tasks with the help of a set of tools for different purposes described in the image from knowledge retrieval to coding to image analysis. Each tool is composed of a set of AI agents that collaborate to solve the query received from the User and return the results to the core agents. Each individual AI agent in AtomAgents is assigned a distinct profile that defines its role and may be powered by a general-purpose LLM from the OpenAI GPT family. The entire process is automated, providing a robust framework for solving challenging tasks in alloy design and analysis with minimal or no human intervention.

Table 1. LLM-powered agents implemented in the current study to solve multiobjective tasks in the context of alloy design and analysis

Agent # (LLM)	Agent name	Agent role
1 (None)	User	Human user that poses the task and can provide feedback at different stages of the problem-solving process.
2 (GPT-4o)	Assistant	Core agent who has access to external tools, including "Planning tool," and provides their input parameters.
3 (GPT-4-turbo)	Scientist	Who is an expert in materials science and can propose research hypotheses.
4 (GPT-4o)	Group chat manager	Chooses the next speaker and broadcasts the message to the whole group.
5 (GPT-4-turbo)	Planner	Suggests a detailed plan to solve the task.
6 (GPT-4-turbo)	Critic	Provides feedback on the plan and revises it when necessary.
7 (GPT-4-turbo)	Engineer	Has access to a library of computational and theoretical tools to solicit new physics data.
8 (GPT-4o)	Coder	Writes Python code.
9 (GPT-4o)	Plot analyzer	Multimodal agent capable of reasoning over visual data including plots.
10 (GPT-4o)	Assistant (knowledge retrieval)	Retrieves knowledge from a document.
11 (GPT-4o)	Reviewer (knowledge retrieval)	Checks the responses from the Assistant for correctness.

agents are powered by a state-of-the-art general-purpose LLM from the GPT family (76) accessed via the OpenAI Application Programming Interface (API) (77). Moreover, each agent is characterized by a unique profile that describes its role in the system. Our team of agents with the entities shown in Table 1, collaborate in a dynamic environment to solve multiobjective tasks in the context of alloy design. The full profile of the agents is shown in *SI Appendix*, Figs. S1–S9.

At the core of AtomAgents lies a group of agents—User, Assistant, Scientist, and Group Manager—who control the overall workflow of the problem-solving process by calling and executing relevant tools, providing appropriate inputs, and returning output results. Several tools are implemented in AtomAgents to facilitate the alloy design and analysis process, performing different tasks as shown in Fig. 2. These tools cover a spectrum of capabilities and functionalities, including computations by atomistic simulations, knowledge retrieval from external sources, coding, plotting, and image analysis, collectively making AtomAgents a robust physics-aware LLM-based framework for solving intricate materials design and analysis tasks. Moreover, as shown in Fig. 2, each tool is equipped with a set of agents that autonomously collaborate to respond to the given query. The full description of the tools and functions incorporated in AtomAgents is listed in *SI Appendix*, Table S1. Notably, the computation tool is composed of a rich library of functions covering a wide range of atomistic simulations such as elastic constant and surface energy calculations, and performing complex nudged elastic band computations. The full list of computation functions along with their profile is listed in *SI Appendix*, Table S2.

A fundamental tool is the planning tool, which consists of a group of agents: an Admin, a Planner, and a Critic, who are responsible for providing a detailed, well-structured plan to solve the multiobjective complex task. When a problem is posed, the planning tool is executed at the beginning, and a plan is created through collaboration between the Planner and the Critic, and returned to the core agents. Then, the core agents start executing the plan by calling the appropriate tools.

The following sections present a series of experiments to demonstrate how the multiagent system addresses various tasks in the domain of alloy design, particularly through generating new physics via atomistic simulations, eliminating or substantially reducing the need for human intervention. All atomistic simulations are performed using LAMMPS (75) at zero temperature. Due to the weak performance of current state-of-the-art LLMs in constructing LAMMPS scripts, all the atomistic simulations are performed by human-generated LAMMPS scripts integrated into the model as python functions and executed via the computation tool. An example of such script is shown in *SI Appendix*, Fig. S10.

The provided experiments highlight different strengths of the developed multiagent system. Specifically, we show that the model can a) integrate materials properties from diverse sources, b) tackle multimodal problems involving image analysis, c) solve multiscale problems connecting microscale feature to macroscopic properties, and generate and validate new hypotheses through atomistic simulations. These complex processes typically demand expertise from multiple domains, sophisticated reasoning abilities, and advanced scripting skills to execute atomistic simulations—tasks that have traditionally required significant manual effort. Our current approach showcases the capability of AI multiagent modeling to not only seamlessly integrate different modalities but also to automate and optimize the entire workflow. By leveraging logic and reasoning capabilities of LLMs, the system can efficiently manage and solve intricate alloy design problems, significantly reducing the need for manual intervention. This innovation not only enhances the precision and effectiveness of alloy design processes but also accelerates discovery and development, paving the way for more advanced and efficient materials engineering.

1.1. Experiment I – Material Properties Calculation and Knowledge Retrieval; Multimodal Integration Problem. Computing material properties such as lattice and elastic constants is crucial in atomistic simulations. These properties characterize materials and are particularly important for theoretical modeling. For instance, when computing the critical fracture toughness based on Griffith's theory, both the elastic constants and surface energies must be determined. However, computing various material properties often requires setting up different structures, writing and adjusting multiple LAMMPS scripts, and running numerous simulations, which becomes cumbersome when studying a broad range of materials, especially in alloy design.

In this section, we demonstrate the capability of multiagent modeling to perform complex tasks without the need for extensive coding knowledge or expertise in LAMMPS. Furthermore, we illustrate how multiagents can automate the extraction of valuable knowledge from the literature, such as material properties computed by specific interatomic potentials from corresponding papers. The entire problem-solving process is managed by AI agents, encompassing plan definition, simulation execution, knowledge retrieval, and the storage of results. The workflow illustrating multiagent collaboration for this experiment is shown in Fig. 3.

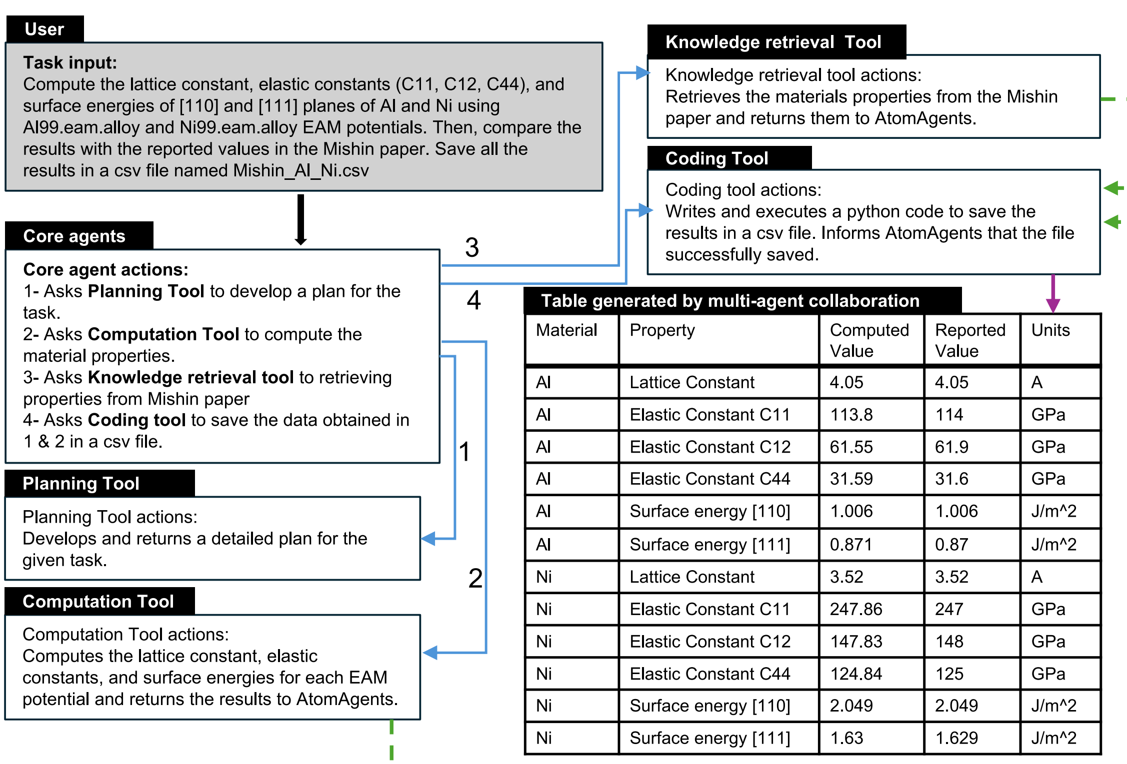
Initially, the core agents activate the "planning" tool responsible for developing a plan to address the query posed by the user. The "planner" agent within this tool breaks down the complex task into simpler subtasks by proposing a step-by-step plan. This plan is then evaluated and approved by the accompanying "critic" agent, and the approved plan is returned to the AtomAgents for execution. The plan involves using a computation tool to derive material properties from atomistic simulations, a knowledge retrieval tool to extract these properties from papers, and coding tools to write a Python script for saving the results.

The "Assistant" agents in AtomAgents demonstrate excellent performance in following the developed plan, calling, and executing the relevant functions. They are particularly adept at providing the correct query to the team of agents embedded in different tools, as these agents do not have access to external information. Once the results from the computation and knowledge retrieval tools are returned, they are collected and formatted as a Python dictionary. This dictionary is then utilized by the "coder" agent to save the results in a CSV file, as illustrated in Fig. 3 in tabular format.

This experiment showcases the efficacy of the multiagent system in solving complex tasks that involve conducting atomistic simulations. Moreover, the agreement between the computed and reported values confirms the accuracy of the computations, paving the way for more challenging experiments.

1.2. Experiment II – Analyzing Screw Dislocation Core Structure; Multimodal Analysis Problem. The previous experiment focused on computing basic material properties through atomistic simulations conducted on pristine materials. In this experiment, we extend the scope to include simulations on defected structures, specifically modeling screw dislocations in body-centered cubic (BCC) materials. Screw dislocations are line defects that significantly influence the plasticity of BCC materials and have been extensively studied through atomistic simulations. A critical aspect of these studies is the dislocation core structure and its interaction with nearby solutes. Many empirical potentials suggest that the screw dislocation core in pure BCC metals is polarized, whereas ab initio simulations based on DFT indicate an unpolarized, compact core structure. The nature of the core structure profoundly affects the movement of long screw dislocations and is crucial for accurately investigating dislocation motion mechanisms, particularly double-kink nucleation. To evaluate the performance of these potentials in predicting the core structure accurately, we analyze differential displacement maps. These maps are generated by subtracting the atomic positions in dislocated structures from those in pristine conditions. Consequently, this experiment employs a specialized multimodal agent equipped with image reasoning capabilities, enhancing the model's performance ability to accurately characterize and understand these complex defects.

In this segment, we address a complex multiobjective problem where a team of agents is tasked with determining the screw dislocation core structure in BCC tungsten using two commonly employed Embedded Atom Method (EAM) (78) potentials: Zhou–Johnson (79) and Marinica (EAM4) (80). Fig. 4 illustrates the workflow executed by the agents to conduct this experiment. The process begins with a query from the user and proceeds through meticulous planning, execution of atomistic simulations, and culminates in image analysis to examine the generated differential displacement (DD) maps. The plan ensures the correct tools are executed with the



Material	Property	Computed Value	Reported Value	Units
Al	Lattice Constant	4.05	4.05	A
Al	Elastic Constant C11	113.8	114	GPa
Al	Elastic Constant C12	61.55	61.9	GPa
Al	Elastic Constant C44	31.59	31.6	GPa
Al	Surface energy [110]	1.006	1.006	J/m^2
Al	Surface energy [111]	0.871	0.87	J/m^2
Ni	Lattice Constant	3.52	3.52	A
Ni	Elastic Constant C11	247.86	247	GPa
Ni	Elastic Constant C12	147.83	148	GPa
Ni	Elastic Constant C44	124.84	125	GPa
Ni	Surface energy [110]	2.049	2.049	J/m^2
Ni	Surface energy [111]	1.63	1.629	J/m^2

Fig. 3. Overview of the multiagent collaboration to solve the complex task posed in Experiment I. After receiving the task from the user, the core agents call the "planning" tool to create a plan for the task. Then the core agents start executing the plan by using "computation" tool to compute the material properties and "knowledge retrieval" tool to retrieve the material properties from a set of scientific papers or other documents. Finally, all the data are collected and sent to "Coding" tool to save them in a comma-separated values (CSV) file.

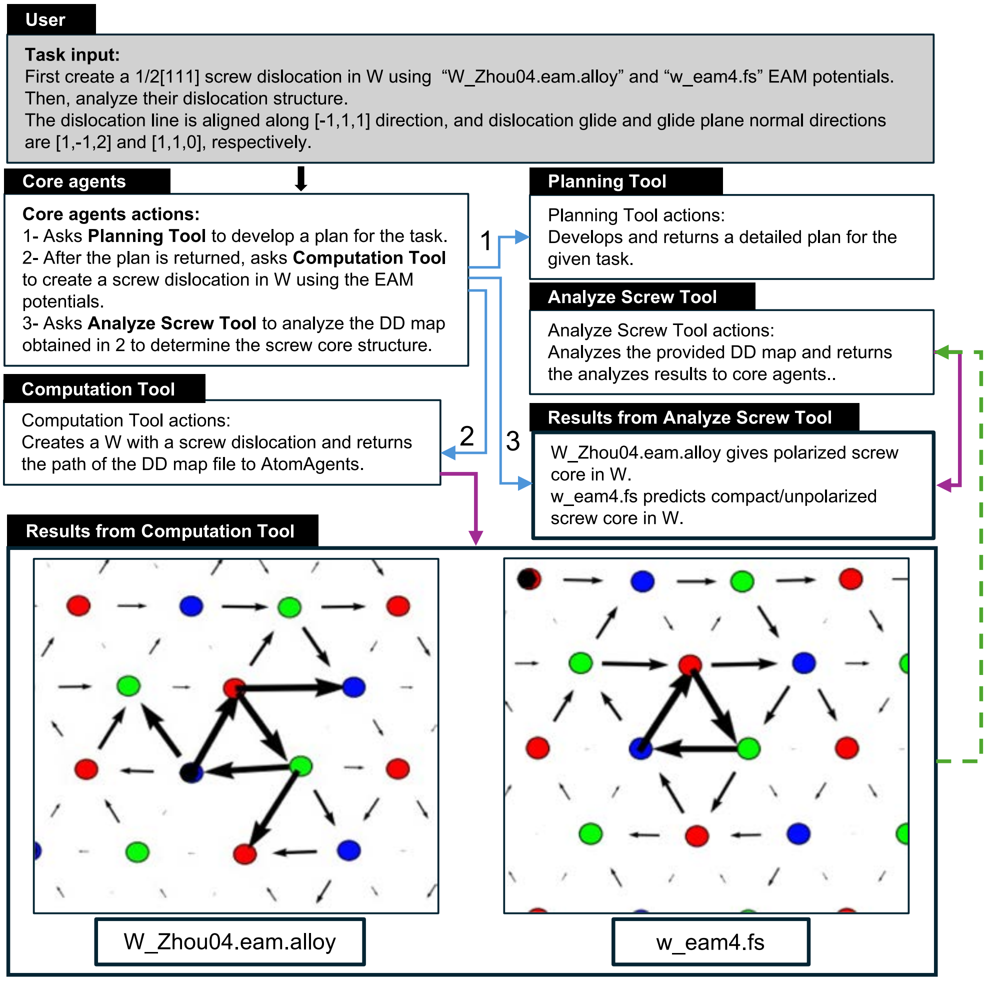


Fig. 4. Overview of the multiagent collaboration to solve the complex task posed in Experiment II. After receiving the task from the user, the core agents call the “planning” tool to create a plan for the task. Then the core agents start executing the plan by using “computation” tool to generate the DD maps and subsequently “Analyze Screw” tool to determine the screw core structure. The DD maps returned by the “computation” tool and the summary of the image analysis made by the “Analyze Screw” tool are included in the image.

appropriate input parameters. Furthermore, the computation tool performs all necessary tasks to create the differential displacement maps without errors. This includes creating the potential file describing pair_style and pair_coeff parameters, computing the lattice constant required to create the BCC structure, creating and relaxing a screw dislocation within tungsten, and constructing the differential displacement maps as shown in Fig. 4.

The last and crucial step in this experiment is the determining of the core structure of the screw dislocations from the generated DD maps. This is done by our multimodal agent capable of reasoning over images. However, our initial assessments indicated the poor performance of the multimodal agent equipped with GPT-4o (or GPT-4v) in distinguishing between the polarized and unpolarized core structures. To circumvent this, we leverage in-context learning to empower the multimodal agent to correctly identify screw core structure by providing it an image including polarized and unpolarized cores with corresponding labels as illustrated in *SI Appendix*, Fig. S11.

The results from the “Analyze Screw Core” tool indicate that the multimodal agent has successfully identified the screw dislocation core structure in W generated by the two EAM potentials. After the successful completion of all the computation and analysis tasks, a summary of the final results is returned by the “Assistant” agent:

The collaborative efforts of the agents in this experiment, leveraging computational techniques and multimodal reasoning capabilities, ensure a thorough and precise evaluation of the dislocation core structures as predicted by the selected potentials.

1.3. Experiment III – Multiscale Mechanics Problem: Fracture Toughness in Alloy Systems. To further explore the capabilities of our multiagent model in solving complex simulation-based problems, this experiment focuses on addressing a multiscale mechanics challenge that combines simulations with theoretical

Box 1.

Assistant:

The analysis of the screw dislocation core structures for the 1/2[111] screw dislocation in Tungsten (W) using two different potentials has been completed:
1. **Using W_Zhou04.eam.alloy Potential:**
 – The screw dislocation core structure is identified as a **polarized/noncompact core**.
 – The displacement vectors around the core form a more complex, nontriangular pattern, indicating a noncompact core structure.
2. **Using w_eam4.fs Potential:**
 – The screw dislocation core structure is identified as a **compact/unpolarized core**.
 – The displacement vectors are directed toward the center of a triangular arrangement, indicating a compact core structure.

models. Specifically, this task involves computing the critical mode I stress intensity factors for Griffith cleavage, K_{Ic}, and dislocation emission, K_{Ie} . The competition between crack cleavage (characteristic of brittle materials) and dislocation emission (characteristic of ductile materials) can be evaluated based on these two parameters. According to linear elastic fracture mechanics (refer to Section 3 for more details), calculating K_{Ic} requires determining the elastic constants and surface energies of the material, while computing K_{Ie}, requires the unstable stacking fault energy for slip along the emission plane in addition to elastic constants. Therefore, this task necessitates the computation of several material properties through atomistic simulations. Furthermore, this experiment extends beyond pure materials, as explored in previous experiments, by incorporating simulations on alloy systems, thus adding another layer of complexity.

The overall workflow employed by our multiagent system is illustrated in Fig. 5, where the user poses a highly complex multi-objective task to explore the effect of Niobium concentration on the fracture toughness for cleavage and dislocation emission in NbMo alloys. An important feature of this experiment is the use of a moment tensor potential (81, 82), a machine-learning-based potential, to describe the interatomic interactions. This highlights the flexibility of the multiagent system in leveraging state-of-the-art machine learning interatomic potentials. The user then asks to explore the effect of Nb concentration on the critical fracture toughness for cleavage, K_{Ic}, and edge dislocation emission, $K_{I}e$ in NbMo alloys for specific crack systems and emission planes, across a range of Nb concentrations (from 0 to 100 in intervals of 20). Additionally, for surface energy and stacking fault energy calculations, a lattice size of 50 Å along free surfaces and 20 Å along other directions is specified. It is worth noting that in alloy systems, fluctuations in random solute environments require sampling multiple configurations to reliably compute average material properties such as lattice constants, elastic constants, surface energies, and stacking fault energies, which increases the computational cost. For this study, we have set the default number of samples to five, but the user can increase this number as needed [e.g., assigning 20 samples for nudged elastic band (NEB) simulations in Experiment IV].

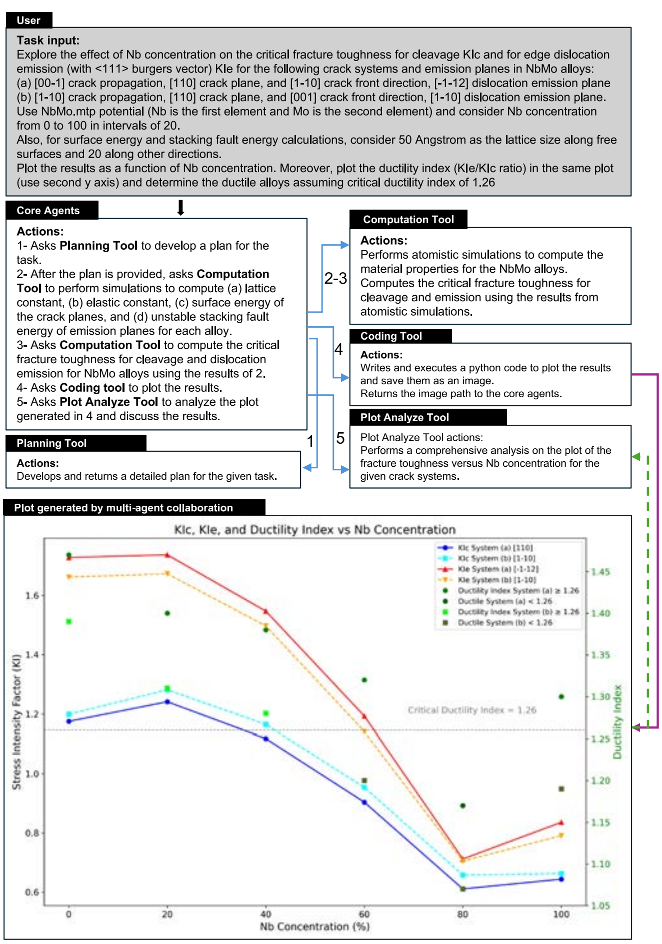


Fig. 5. Overview of the multiagent collaboration to solve the complex task posed in Experiment III. After receiving the task from the user, the core agents call the "planning" tool to create a plan for the task. Then the core agents start executing the plan by using "computation" tool to compute the elastic constants and surface energies, and eventually the fracture toughness. Subsequently "Coding" tool is used to plot the results which are then analyzed by the "Plot analyze" tool via a multimodal agent.

The user also requests plotting K_{Ic} and K_{Ie} as a function of Nb concentration. Additionally, the user asks to plot the ductility index, defined as $D = K_{Ie}/K_{Ic}$, using a secondary y-axis and to identify ductile alloys based on a critical ductility index value of $D_c = 1.26$. Alloys with a ductility index below D_c are categorized as ductile, while those above are considered brittle. The chosen D_c value ensures alignment with real experimental observations at room temperature, correctly predicting Mo and W as brittle and Nb, Ta, and V as ductile (16, 83).

At the initial step, the core agents engage the "planning" tool to provide a detailed implementation plan for the given task. The "planner" agent develops a comprehensive plan, which is then validated for accuracy by the "critic" agent. The plan (shown in *SI Appendix*, Fig. S12), demonstrates the planner's effectiveness in identifying the key steps necessary for accomplishing the tasks posed by the user. Specifically, the planner correctly identifies that calculating k_{Ic} requires computing the surface energy and elastic constants, and stacking fault energy is needed for K_{Ie}. Moreover, it accurately recognizes that the surface energy should be computed along the crack planes and stacking fault plane should align with the dislocation emission plane. Collecting the final results, plotting, and analyzing them to identify the ductile alloys is another crucial step adeptly captured in the plan. However, we noticed that the agent suggests alloys with a ductility index above 1.26 as ductile, which is incorrect. The multiagent system allows users to intervene at various stages, so we will address this issue once all the results have been collected by reminding the agent that a ductile alloy requires $D < D_c$.

Upon receiving the plan, the "Assistant" agent in the core module initiates the implementation of the plan step-by-step by activating the tools as outlined. Despite the complexity and length of the tasks, all tools are implemented and the necessary results are gathered autonomously. After the material properties are collected, the fracture toughness for crack cleavage and dislocation emission are computed for the specified crack systems and alloy compositions. These results, along with their ratios, are then used as input for the plotting tool.

The plot of the results, generated by the multiagent collaboration, is displayed in Fig. 5 Subsequently, the plot is analyzed, and conclusions are drawn by the multimodal agent within the analysis tool, as shown in *SI Appendix*, Fig. S13. As highlighted by the multimodal agent, both systems exhibit increased ductility at higher Nb concentrations. Crack System B, with a [1-10] dislocation emission plane, shows the lowest ductility index, indicating that NbMo alloys with an Nb concentration greater than 60% are ductile.

The findings offer a detailed analysis of crack behavior in BCC alloys and demonstrate the multimodal agent's capability to provide a robust framework for automating alloy design. By integrating various modalities such as theory, simulations, coding, and image analysis, the multiagent system showcases its comprehensive capabilities.

1.4. Experiment IV – Hypothesis Generation and Validation. The previous experiments demonstrated the exceptional performance of our multiagent model in solving complex computational problems requiring atomistic simulations. In this experiment, we explore another intriguing aspect of the multiagent strategy: hypothesis generation and validation. Due to their high reasoning capabilities, general-purpose LLMs like the GPT family have the potential to generate innovative research ideas across fields, including materials science (84). Combining this with the computational prowess of the multiagent strategy offers a robust framework where LLM-based research ideas and hypotheses can be validated and refined, providing opportunities for scientific discovery. The objective of this experiment is to assess the performance of our multiagent model in hypothesis generation and validation.

It is well known that plastic deformation in BCC metals is dominated by <111> screw dislocation motion. An important physical property of BCC materials is the Peierls potential, which is the energy change of the straight dislocation line as it moves from one Peierls valley (minimum energy configuration) to the next. The maximum energy along this profile is the Peierls barrier, which profoundly affects screw dislocation motion and, consequently, the mechanical properties of BCC materials. Therefore, accurately estimating the Peierls potential is essential for predicting and understanding the behavior of BCC materials. The Peierls barrier can be computed using transition state calculations such as the NEB method. While the Peierls potential in pure BCC metals is constant and has been extensively studied (21, 85–89), the energy landscape in alloys is more complex, leading to a wide distribution of energy barriers. Consequently, multiple NEB simulations are required to accurately estimate the mean barrier. However, this process becomes computationally expensive when screening the entire design space, for instance, to search for compositions with high energy barriers. Therefore, finding a correlation between the Peierls barriers and more easily computed properties could significantly save time by narrowing down the composition space toward designs with higher barriers and thus enhanced mechanical performance.

To facilitate hypothesis generation, a "Scientist" agent is integrated into our model, profiled to act as a leading materials scientist. In the current framework, the user can directly collaborate with the Scientist to generate innovative hypotheses. However, it is possible to integrate other mechanisms for hypothesis generation, such as two AI agents that adversarially collaborate to push each other toward more innovative and robust research ideas, or extracting and generating novel research ideas from powerful knowledge graphs (67, 70, 90). Fig. 6 shows a summary of the hypotheses generated by the "Scientist" agent in response to the user's query. It can be observed that, the scientist proposes a positive correlation between the Peierls barrier and the SD of the energy change. The proposed correlation by the "Scientist" is particularly encouraging as it reduces the computational burden of performing costly NEB simulations to compute the Peierls barrier directly.

We now utilize our multiagent system to test the hypotheses generated in the previous experiments using atomistic simulations. As in the earlier experiment, we use NbMo as our model

Chat between a User and AI scientist

User

```
In random alloys, dislocations at different minimum
locations have different potential energies.  Therefore,
when a dislocation moves from a minimum position to
the next one, the net energy change might not be zero.
However, for a large number of dislocations, the average
of the net energy change might be zero with a non-zero
standard deviation.  Do you think there is a correlation
between the standard deviation of the energy changes and
the Peierls barrier, which is the maximum energy between
two adjacent energy minima?
```

Scientist (GPT-4-turbo)

```
Here are the summarized hypotheses regarding the
correlation between the standard deviation of energy
changes in dislocation movements and the Peierls barrier
in random alloys:

1.  **Correlation with Peierls Barrier**:  A higher
standard deviation in the energy changes suggests a
more irregular energy landscape, potentially leading to
higher Peierls barriers.  This implies that materials
with greater variability in their energy landscapes
might generally exhibit higher Peierls barriers.

2.  **Effect on Dislocation Mobility**:  Increased
variability (higher standard deviation) in the energy
landscape can reduce dislocation mobility due to higher
and more variable Peierls barriers, impacting the
material's mechanical properties such as strength and
ductility.

3.  **Material Design Implications**:  By understanding
and manipulating the standard deviation of energy
changes, it may be possible to tailor the Peierls
barrier and, consequently, the mechanical properties of
alloys to meet specific performance criteria.

While a direct quantitative correlation would require
detailed computational or experimental studies specific
to the alloy system in question, qualitatively, a higher
standard deviation in energy changes among dislocations
is likely indicative of a higher and more variable
Peierls barrier, affecting dislocation mobility and
material properties.
```

Fig. 6. The hypotheses generated by the "scientist" agent empowered by GPT-4-turbo in response to the user's query.

system, described by the moment tensor potential, though it is feasible to apply other alloy systems and interatomic potentials. The overall workflow of the problem-solving process is depicted in Fig. 7, where the user assigns a computational task pertaining to the aforementioned hypotheses, followed by a detailed plan from the "Planner." Subsequent computations are performed by the "Computation Tool" to compute the energy change and Peierls barrier from NEB simulations. The details of these NEB simulations can be found in Section 3.

After the completion of the NEB simulations, the results are collected and sent to the "Coding Tool" for plotting. Fig. 7 presents the plot of the Peierls barrier mean and SD against the potential energy change SD. The analysis of this plot is shown in *SI Appendix*, Fig. S14. The comprehensive analysis reveals the proficient performance of the multimodal agent, empowered by GPT-4o, in analyzing the plot. Following a thorough review, the agent concludes that there is a positive correlation between the mean Peierls barrier and the SD of the potential energy change, hence verifying the hypothesis developed by the "Scientist." This finding is significant as it points toward a method for accelerating the design process of

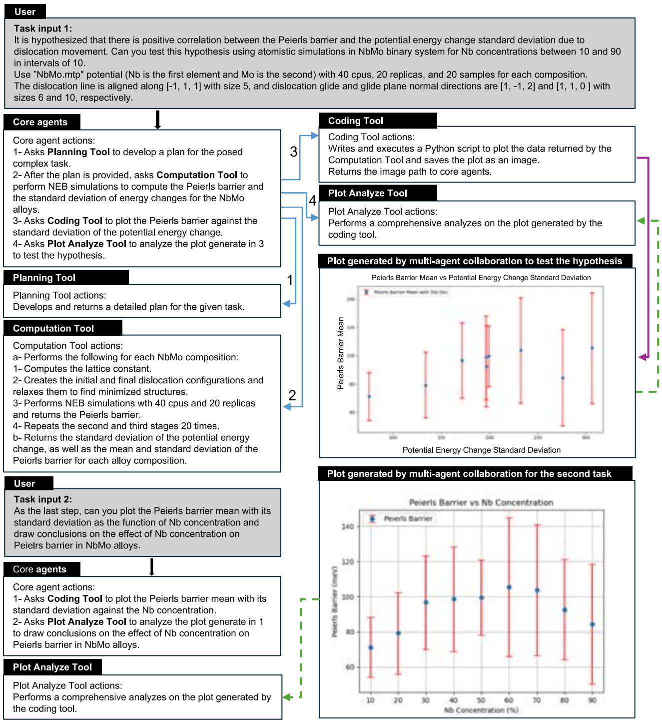


Fig. 7. Overview of the multiagent collaboration to solve the complex task posed in Experiment IV and then a follow-up task. After receiving the task from the user, the core agents call the "planning" tool to create a plan for the task. Then the core agents start executing the plan by using "computation" tool to compute the Peierls barrier and the SD of the potential energy changes for a set of alloys. Subsequently "Coding" tool is used to plot the results which are then analyzed by the "Plot analyze" tool via a multimodal agent. In the second task, the results of the first task are plotted via the coding agent to show the variation of the Peierls barrier against Nb concentration. The final analysis from the multiagent model finalizes the task.

materials with enhanced mechanical performance by focusing on specific energy characteristics that are indicative of higher barriers.

AtomAgents facilitates human–AI collaboration at various stages of the problem-solving process. For example, after the multiagent system successfully completed the complex computations and analyses required to test the hypotheses, we posed another task to explore the effect of Niobium concentration on the Peierls barrier in NbMo alloys. Since the necessary data had already been collected during the main task, no further simulations were required. Fig. 7 displays the follow-up task posed by the user to the multiagent system, along with the plot generated by the "Coding Tool." This plot is subsequently evaluated by the "Analyze Plot" tool to draw conclusions about the impact of the Nb concentration on the Peierls barrier in NbMo alloys. The results of this analysis are depicted in *SI Appendix*, Fig. S15, demonstrating the proficiency of the multimodal agent, powered by GPT-4o, in analyzing the variations of the Peierls barrier with respect to the Nb concentration. This capability highlights the system's effectiveness in leveraging existing data to generate insights without the need for additional computational resources, thus streamlining the research process and enhancing the efficiency of materials design evaluations.

2. Summary and Future Perspective

We combined the remarkable capabilities in complex reasoning, rational thinking, and strategic planning of LLMs with abilities to write and execute code, run atomistic-level molecular simulations to solicit new data, and conduct visual analysis of graphed data to develop key insights that form critical steps to form complex problems. This agentic strategy allowed us to build physics-aware intelligent systems capable of solving complex analysis and design tasks.

We showed that such an integration is possible to model complex compositional problems in alloy modeling and design. Our model, AtomAgents, combines a comprehensive suite of tools with various capabilities, including physics-based atomistic simulators, effectively coupling the formidable multimodal reasoning capabilities of LLMs with the essential computational power required for addressing sophisticated materials design problems. The design of AtomAgents prioritizes flexibility, making it easy to add and customize tools for a wide variety of atomistic simulations and computational tasks. This adaptability ensures that AtomAgents is not restricted to a narrow set of problems but can evolve to tackle new challenges as they arise.

To mitigate risks associated with LLMs, AtomAgents' tools are designed with predefined key parameters and integrated checks to ensure alignment with user requests and adaptability to emerging behaviors. Moreover, our AI model incorporates a human-in-the-loop design that fosters dynamic collaboration between humans and AI, allowing real-time integration of expert knowledge. By actively involving human experts, we ensure effective oversight and guidance, reducing the risk of failures such as hallucinations or erroneous predictions, and enhancing the reliability of our AI models predictions.

A key feature of the proposed multimodal system is the potential of integration of different modalities from different domains into the materials design process, offering a paradigm shift in solving challenging multiscale materials problems that traditionally demands considerable human expertise and manual effort. The comprehensive data integration allows AtomAgents to continuously update and improve its predictive capabilities, ensuring that the latest findings and trends in materials science are incorporated into its models. The systematic assimilation of such diverse and dynamic data sources not only enhances the accuracy of simulations and predictions but also accelerates the materials discovery process, paving the way for groundbreaking advancements in the field.

While the research reported here primarily focuses on designing binary alloys and metallic materials, the AtomAgents framework inherently offers the flexibility to explore more complex systems, such as high-entropy alloys (91, 92). These systems open a vast compositional design space, providing unprecedented opportunities to engineer materials with novel properties like enhanced strength and fracture toughness (93, 94). However, navigating this extensive space presents significant challenges, requiring innovative and precise methodologies to manage its complexities effectively. AtomAgents, with its multiagent architecture, is ideally equipped for these tasks, utilizing collective intelligence to efficiently explore and exploit this expansive design space. Furthermore, our framework is not confined to just metallic materials; owing to the versatility of LAMMPS, which serves as our physics engine, AtomAgents can also model a wide range of other materials systems-including biomaterials, polymers, ceramics, and liquids, that can be readily simulated within the LAMMPS environment.

However, to fully leverage this flexibility and accommodate a diverse range of materials systems, efficient and automated script generation remains a critical challenge. Despite significant advancements in LLMs, these models are not yet fully capable of independently generating LAMMPS scripts for complex materials simulations, particularly those involving intricate defect structures. Nevertheless, with continued progress in LLM technology and the potential for fine-tuning these models specifically for materials applications, this limitation may be addressed in the near future. Moreover, we believe that the multiagent system has the potential to alleviate this challenge by autonomously managing and generating the necessary scripts. However, fully developing and integrating this capability is beyond the scope of the current paper and remains a topic for future work.

In the meantime, our strategy involves predefining key computational functions to streamline workflows and accelerate alloy design efforts. This approach reduces the need for generating new scripts for each individual material or simulation while maintaining essential flexibility. Additionally, the modular and adaptable design of AtomAgents allows tools and scripts to be customized for various design tasks and material types. This balanced strategy ultimately enables AtomAgents to efficiently handle varying computational demands and the complexities associated with different material-specific challenges.

The potential of integrating advanced deep learning models and generative tools within AtomAgents offers a promising avenue to enhance its capabilities. These AI-driven models can predict material properties across diverse alloy systems, potentially reducing reliance on extensive atomistic simulations (95, 96). Such integration could significantly streamline the process, reducing both time and computational overhead, especially when dealing with complex multicomponent alloys (97). This strategic enhancement would not only optimize the efficiency of material design but also expand the boundaries of what can be achieved in materials science, propelling AtomAgents to the forefront of innovative materials discovery.

Working as the integrative agent in the multiagent system, LLMs play a crucial role in orchestrating the interactions and operations within these systems. They facilitate critical steps such as planning, reasoning, and critical thinking, thereby shaping the system's overall efficiency and output. Therefore, the accuracy and performance of the entire model heavily depend on the capabilities of the underlying LLM. As advancements in LLMs continue, their enhanced computational power and refined algorithms can significantly boost the performance of multiagent systems. Moreover, the modular architecture of our approach allows for the integration of different LLMs tailored to the specific needs of individual agents. The integration of high-performance, open-source foundation models such as Llama 3 and Mistral/Mixtral, alongside specialized smaller models like Phi-3 or fine-tuned models in scientific subjects (62), opens up exciting possibilities for further enhancement (98). As these foundation models continue to advance in capability, the potential for our multiagent systems to improve becomes even more pronounced. This adaptability not only ensures that our systems stay at the cutting edge of technological advancements but also signifies that the effectiveness and efficiency of AtomAgents will invariably increase. By continuously incorporating more capable LLMs, our systems can achieve deeper insights and more precise predictions, leading to accelerated innovation in materials design and discovery.

3. Materials and Methods

3.1. Bulk Crystals Generation. We used Atomic Simulation Environment (99) to generate bulk crystal structures.

3.2. Screw Dislocation Generation. To model the screw dislocation, a periodic array of dislocations (PAD) configuration is used (e.g. ref. 100) with periodic boundary conditions along the dislocation glide direction z||[11$\bar{2}$] and dislocation line direction x||[111], and free boundaries along the glide plane normal direction y||[$\bar{1}$10]. Atomic positions are relaxed by using a combination of the FIRE algorithm (101) and relaxation of the cell dimensions until the convergence is achieved–the norm of the force vector fell below 10^{-6} eV/Å and stresses σ_{xx}, σ_{xy}, and σ_{yy} fell below 0.1 MPa.

3.3. NEB Simulations. To compute the minimum energy path between initial and final screw dislocation configurations, NEB (102–104) computations are performed as implemented in LAMMPS. To perform NEB simulations, first, the initial and final screw dislocation configurations were created inside the material using the PAD method. Then NEB simulations are performed using the desired number of replicas. An initial path of intermediate configurations (replicas) is constructed by linearly interpolating the atomic positions between the relaxed initial and final states. The NEB interreplica spring constant is set to 10^{-2} eV/Å^2 and convergence is assumed when the maximum of the force acting on all of the atoms across all replicas is less than 10^{-3} eV/Å.

3.4. Surface Energy Calculations. The surface energies along a given plane are computed as follows. First, system with periodic boundary conditions in all directions is created and relaxed and the energy E^{bulk} is computed. Then, the supercell is extended along the desired direction and then the system is relaxed and the energy is computed E^{surf}. The surface energy is then computed as

$$\gamma_s = \frac{E^{\mathrm{surf}} - E^{\mathrm{bulk}}}{2A},$$

where A is the area of the surface plane.

3.5. Critical Fracture Toughness for Cleavage Fracture and Dislocation Emission. The mode I anisotopic critical fracture toughness for cleavage fracture and dislocation emission is given by (e.g. refs. 16, 105, and 106.)

$$K_{Ic} = \sqrt{\frac{2\gamma_s}{\lambda_{22}(C)}}, \quad K_{Ie} = \frac{\sqrt{\gamma_{\mathrm{usf}} o(C, \theta, \phi)}}{F_{12}(C, \theta)} \cos(\phi),$$

where γ_s is the surface energy, γ_{usf} is the unstable stacking fault energy, and C is the material's elastic tensor.

3.6. Agent Design. We design AI agents using the state-of-the-art all-purpose LLM GPT-4 family models. The dynamic multiagent collaboration is implemented in the AutoGen framework (107), an open-source ecosystem for agent-based AI modeling. In our multiagent system, the user, admin, and executor agents are constructed using UserProxyAgent class from Autogen, and assistant and reviewer agents in the "Knowledge retrieval" tool are crated using RetrieveAssistantAgent class. Moreover, the "Plot analyzer" agent is created via the MultimodelConversableAgent class. The remaining agents are created via AssistantAgent class from Autogen; and the group chat manager is created using GroupChatManager class. Each agent is assigned a role through a profile description, included as *system_message* at their creation. Details of the implementation can be identified in the source code shared via GitHub.

3.7. Function and Tool Design. All the tools implemented in this work are defined as Python functions. Each function is characterized by a name, a description, and input properties with a description.

Data, Materials, and Software Availability. GitHub repo data and codes have been deposited in https://github.com/lamm-mit/AtomAgents (108). Alternatively, they will be provided by the corresponding author based on reasonable request. All other data are included in the manuscript and/or *SI Appendix*.

ACKNOWLEDGMENTS. We acknowledge support from the United States Department of Agriculture (2021-69012- 35978), Department of Energy - Strategic Environmental Research and Development Program (WP22-S1-3475), Army Research Office (79058LSCSB, W911NF-22-2-0213, and W911NF2120130) as well as the Massachusetts Institute of Technology - IBM Watson AI Lab, Massachusetts Institute of Technology's Generative AI Initiative, and Google. Additional support from the National Institutes of Health (U01EB014976 and R01AR0 77793) is acknowledged. A.G. gratefully acknowledges the financial support from the Swiss National Science Foundation (project #P500PT_214448). A.G. thanks Predrag Andric from Svenska Kullagerfabriken for providing the Python functions used to compute the critical fracture toughness for dislocation emission.

1. D. B. Miracle *et al*., Exploration and development of high entropy alloys for structural applications. *Entropy* **16**, 494–525 (2014).
2. C. Wang *et al*., Advanced carbon for flexible and wearable electronics. *Adv. Mater.* **31**, 1801072 (2019).
3. D. Raabe, The materials science behind sustainable metals and alloys. *Chem. Rev.* **123**, 2436–2608 (2023).
4. S. C. Shen *et al*., Computational design and manufacturing of sustainable materials through first-principles and materiomics. *Chem. Rev.* **123**, 2242–2275 (2023).
5. J. A. Rogers, T. Someya, Y. Huang, Materials and mechanics for stretchable electronics. *Science* **327**, 1603–1607 (2010).
6. X. Zhang, Y. Chen, J. Hu, Recent advances in the development of aerospace materials. *Prog. Aerosp. Sci.* **97**, 22–34 (2018).
7. B. Blakey-Milner *et al*., Metal additive manufacturing in aerospace: A review. *Mater. Des.* **209**, 110008 (2021).
8. C. Liu, F. Li, L. P. Ma, H. M. Cheng, Advanced materials for energy storage. *Adv. Mater.* **22**, E28–E62 (2010).
9. S. Koohi-Fayegh, M. A. Rosen, A review of energy storage types, applications and recent developments. *J. Energy Storage* **27**, 101047 (2020).
10. S. Lin, J. Liu, X. Liu, X. Zhao, Muscle-like fatigue-resistant hydrogels by mechanical training. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 10244–10249 (2019).
11. A. Chaves *et al*., Bandgap engineering of two-dimensional semiconductor materials. *NPJ 2D Mater. Appl.* **4**, 29 (2020).
12. S. Suresh, *Fatigue of Materials* (Cambridge University Press, 1998).
13. M. J. Buehler, *Atomistic Modeling of Materials Failure* (Springer, 2008).
14. Z. Wu, W. Curtin, Mechanism and energetics of $< c + a >$ dislocation cross-slip in hcp metals. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11137–11142 (2016).
15. Z. Wu, R. Ahmad, B. Yin, S. Sandlöbes, W. Curtin, Mechanistic origin and prediction of enhanced ductility in magnesium alloys. *Science* **359**, 447–452 (2018).
16. E. Mak, B. Yin, W. Curtin, A ductility criterion for bcc high entropy alloys. *J. Mech. Phys. Solids* **152**, 104389 (2021).
17. M. J. Buehler, F. F. Abraham, H. Gao, Hyperelasticity governs dynamic fracture at a critical length scale. *Nature* **426**, 141–146 (2003).
18. M. J. Buehler, A. Hartmaier, H. Gao, M. Duchaineau, F. F. Abraham, Atomic plasticity: Description and analysis of a one-billion atom simulation of ductile materials failure. *Comput. Methods Appl. Mech. Eng.* **193**, 5257–5282 (2004).
19. R. J. M. Pellenq *et al*., A realistic molecular model of cement hydrates. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 16102–16107 (2009).
20. S. M. Taheri Mousavi, H. Zhou, G. Zou, H. Gao, Transition from source-to stress-controlled plasticity in nanotwinned materials below a softening temperature. *NPJ Comput. Mater.* **5**, 2 (2019).
21. L. Ventelon, F. Willaime, E. Clouet, D. Rodney, Ab initio investigation of the Peierls potential of screw dislocations in bcc Fe and W. *Acta Mater.* **61**, 3973–3985 (2013).
22. M. J. Buehler, A. Hartmaier, H. Gao, Atomistic and continuum studies of crack-like diffusion wedges and associated dislocation mechanisms in thin films on substrates. *J. Mech. Phys. Solids* **51**, 2105–2125 (2003).
23. M. J. Buehler, A. Hartmaier, H. Gao, Hierarchical multi-scale modelling of plasticity of submicron thin metal films. *Model. Simul. Mater. Sci. Eng.* **12**, S391 (2004).
24. A. Hartmaier, M. J. Buehler, H. Gao, Multiscale modeling of deformation in polycrystalline thin metal films on substrates. *Adv. Eng. Mater.* **7**, 165–169 (2005).
25. A. Ghafarollahi, W. A. Curtin, Theory of double-kink nucleation in dilute bcc alloys. *Acta Mater.* **196**, 635–650 (2020).
26. A. Ghafarollahi, W. Curtin, Theory of kink migration in dilute bcc alloys. *Acta Mater.* **215**, 117078 (2021).
27. R. Kubilay, A. Ghafarollahi, F. Maresca, W. Curtin, High energy barriers for edge dislocation motion in body-centered cubic high entropy alloys. *NPJ Comput. Mater.* **7**, 112 (2021).
28. A. Ghafarollahi, W. A. Curtin, Screw-controlled strength of bcc non-dilute and high-entropy alloys. *Acta Mater.* **226**, 117617 (2022).
29. F. Maresca, W. A. Curtin, Theory of screw dislocation strengthening in random bcc alloys from dilute to "high-entropy" alloys. *Acta Mater.* **182**, 144–162 (2020).
30. C. Varvenne, G. P. M. Leyson, M. Ghazisaeidi, W. A. Curtin, Solute strengthening in random alloys. *Acta Mater.* **124**, 660–683 (2017).
31. M. Mozaffar *et al*., Deep learning predicts path-dependent plasticity. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 26414–26420 (2019).
32. X. Liu, C. E. Athanasiou, N. P. Padture, B. W. Sheldon, H. Gao, Knowledge extraction and transfer in data-driven fracture mechanics. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2104765118 (2021).
33. K. Guo, Z. Yang, C. H. Yu, M. J. Buehler, Artificial intelligence and machine learning in design of mechanical materials. *Mater. Horiz.* **8**, 1153–1172 (2021).
34. E. L. Buehler, M. J. Buehler, End-to-end prediction of multimaterial stress fields and fracture patterns using cycle-consistent adversarial and transformer neural networks. *Biomed. Eng. Adv.* **4**, 100038 (2022).
35. Z. Yang, M. J. Buehler, Linking atomic structural defects to mesoscale properties in crystalline solids using graph neural networks. *NPJ Comput. Mater.* **8**, 198 (2022).
36. M. J. Buehler, Modeling atomistic dynamic fracture mechanisms using a progressive transformer diffusion model. *J. Appl. Mech.* **89**, 121009 (2022).
37. A. Merchant *et al*., Scaling deep learning for materials discovery. *Nature* **624**, 80–85 (2023).
38. F. Y. Liu, B. Ni, M. J. Buehler, Presto: Rapid protein mechanical strength prediction with an end-to-end deep learning model. *Extreme Mech. Lett.* **55**, 101803 (2022).
39. M. J. Buehler, A computational building block approach towards multiscale architected materials analysis and design with application to hierarchical metal metamaterials. *Model. Simul. Mater. Sci. Eng.* **31**, 054001 (2023).
40. M. J. Buehler, MeLM, a generative pretrained language modeling framework that solves forward and inverse mechanics problems. *J. Mech. Phys. Solids* **181**, 105454 (2023).
41. A. J. Lew, M. J. Buehler, Single-shot forward and inverse hierarchical architected materials design for nonlinear mechanical properties using an attention-diffusion model. *Mater. Today* **64**, 10–20 (2023).
42. W. Lu, N. A. Lee, M. J. Buehler, Modeling and design of heterogeneous hierarchical bioinspired spider web structures using deep learning and additive manufacturing. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2305273120 (2023).
43. B. Ni, D. L. Kaplan, M. J. Buehler, ForceGen: End-to-end de novo protein generation based on nonlinear mechanical unfolding responses using a language diffusion model. *Sci. Adv.* **10**, eadl4000 (2024).
44. W. Lu, D. L. Kaplan, M. J. Buehler, Generative modeling, design, and analysis of spider silk protein sequences for enhanced mechanical properties. *Adv. Funct. Mater.* **34**, 2311324 (2024).
45. J. J. de Pablo, W. A. Curtin, Multiscale modeling in advanced materials research: Challenges, novel methods, and emerging applications. *MRS Bull.* **32**, 905–911 (2007).
46. F. Roters *et al*., Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: Theory, experiments, applications. *Acta Mater.* **58**, 1152–1211 (2010).
47. J. Fish, G. J. Wagner, S. Keten, Mesoscopic and multiscale modelling in materials. *Nat. Mater.* **20**, 774–786 (2021).
48. A. Jain *et al*., Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
49. M. De Jong *et al*., Charting the complete elastic properties of inorganic crystalline compounds. *Sci. Data* **2**, 1–13 (2015).
50. P. M. Anderson, J. P. Hirth, J. Lothe, *Theory of Dislocations* (Cambridge University Press, 2017).
51. D. Lee, W. Chen, L. Wang, Y. C. Chan, W. Chen, Data-driven design for metamaterials and multiscale systems: A review. *Adv. Mater.* **36**, 2305254 (2024).
52. J. Behler, M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
53. R. Kobayashi, D. Giofré, T. Junge, M. Ceriotti, W. A. Curtin, Neural network potential for Al-Mg-Si alloys. *Phys. Rev. Mater.* **1**, 053604 (2017).
54. T. Mueller, A. Hernandez, C. Wang, Machine learning for interatomic potential models. *J. Chem. Phys.* **152**, 050902 (2020).
55. Y. Zuo *et al*., Performance and cost assessment of machine learning interatomic potentials. *J. Phys. Chem. A* **124**, 731–745 (2020).
56. Y. Mishin, Machine-learning interatomic potentials for materials science. *Acta Mater.* **214**, 116980 (2021).
57. J. Wei *et al*., Emergent abilities of large language models. arXiv [Preprint] (2022). https://doi.org/10.48550/arXiv.2206.07682 (Accessed 29 October 2024).
58. Y. Chang *et al*., A survey on evaluation of large language models. *ACM Trans. Intell. Syst. Technol.* **15**, 1–45 (2024).
59. J. Boyko *et al*., An interdisciplinary outlook on large language models for scientific research. arXiv [Preprint] (2023). https://doi.org/10.48550/arXiv.2311.04929 (Accessed 29 October 2024).
60. Y. Hu, M. J. Buehler, Deep language models for interpretative and predictive materials science. *APL Mach. Learn.* **1**, 010901 (2023).
61. K. M. Jablonka *et al*., 14 examples of how LLMs can transform materials science and chemistry: A reflection on a large language model hackathon. *Digit. Discov.* **2**, 1233–1250 (2023).
62. R. K. Luu, M. J. Buehler, BioinspiredLLM: Conversational large language model for the mechanics of biological and bio-inspired materials. *Adv. Sci.* **11**, 2306724 (2024).
63. K. M. Jablonka, P. Schwaller, A. Ortega-Guerrero, B. Smit, Leveraging large language models for predictive chemistry. *Nat. Mach. Intell.* **6**, 161–169 (2024).
64. M. J. Buehler, MechGPT, a language-based strategy for mechanics and materials modeling that connects knowledge across scales, disciplines, and modalities. *Appl. Mech. Rev.* **76**, 021001 (2024).
65. A. Madani *et al*., ProGen: Language modeling for protein generation. arXiv [Preprint] (2020). https://doi.org/10.48550/arXiv.2004.03497 (Accessed 29 October 2024).
66. A. Vaswani *et al*., Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30**, 6000–6010 (2017).
67. M. J. Buehler, Accelerating scientific discovery with generative knowledge extraction, graph-based representation, and multimodal intelligent graph reasoning. arXiv [Preprint] (2024). https://doi.org/10.48550/arXiv.2403.11996 (Accessed 29 October 2024).
68. M. J. Buehler, Cephalo: Multi-modal vision-language models for bio-inspired materials analysis and design. arXiv [Preprint] (2024). https://doi.org/10.48550/arXiv.2405.19076 (Accessed 29 October 2024).
69. B. Ni, M. J. Buehler, Mechagents: Large language model multi-agent collaborations can solve mechanics problems, generate new data, and integrate knowledge. *Extreme Mech. Lett.* **67**, 102131 (2024).
70. M. J. Buehler, Generative retrieval-augmented ontologic graph and multiagent strategies for interpretive large language model-based materials design. *ACS Eng. Au* **4**, 241–277 (2024).
71. A. Ghafarollahi, M. J. Buehler, ProtAgents: Protein discovery via large language model multi-agent collaborations combining physics and machine learning. *Digit. Discov.* **3**, 1389–1409 (2024).
72. A. M. Bran *et al*., Augmenting large language models with chemistry tools. *Nat. Mach. Intell.* **6**, 525–535 (2024).

73. I. Stewart, M. Buehler, Molecular analysis and design using multimodal generative artificial intelligence via multi-agent modeling. ChemRxiv [Preprint] (2024). 10.26434/chemrxiv-2024-nwm7n (Accessed 29 October 2024).
74. Z. Xi *et al.*, The rise and potential of large language model based agents: A survey. arXiv [Preprint] (2023). https://arxiv.org/abs/2309.07864 (Accessed 20 October 2024).
75. S. Plimpton, Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
76. J. Achiam *et al.*, Gpt-4 technical report. arXiv [Preprint] (2023). https://arxiv.org/abs/2303.0877 (Accessed 29 October 2024).
77. OpenAI, OpenAI API. https://openai.com/blog/openai-api. Accessed 29 October 2024.
78. M. S. Daw, M. I. Baskes, Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Phys. Rev. B* **29**, 6443 (1984).
79. X. Zhou, R. Johnson, H. Wadley, Misfit-energy-increasing dislocations in vapor-deposited CoFe/NiFe multilayers. *Phys. Rev. B* **69**, 144113 (2004).
80. M. C. Marinica *et al.*, Interatomic potentials for modelling radiation defects and dislocations in tungsten. *J. Phys. Condens. Matter* **25**, 395502 (2013).
81. A. V. Shapeev, Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Model. Simul.* **14**, 1153–1173 (2016).
82. S. Yin *et al.*, Atomistic simulations of dislocation mobility in refractory high-entropy alloys and the effect of chemical short-range order. *Nat. Commun.* **12**, 4873 (2021).
83. I. Novikov, O. Kovalyova, A. Shapeev, M. Hodapp, AI-accelerated materials informatics method for the discovery of ductile alloys. *J. Mater. Res.* **37**, 3491–3504 (2022).
84. A. Ghafarollahi, M. J. Buehler, SciAgents: Automating scientific discovery through bioinspired multi-agent intelligent graph reasoning. *Adv. Mater.* 2413523 (2024).
85. L. Proville, D. Rodney, M. C. Marinica, Quantum effect on thermally activated glide of dislocations. *Nat. Mater.* **11**, 845–849 (2012).
86. R. Gröger, V. Vitek, Stress dependence of the Peierls barrier of $1/2 < 111 >$ screw dislocations in BCC metals. *Acta Mater.* **61**, 6362–6371 (2013).
87. L. Dezerald *et al.*, Ab initio modeling of the two-dimensional energy landscape of screw dislocations in BCC transition metals. *Phys. Rev. B* **89**, 024104 (2014).
88. D. Rodney, L. Ventelon, E. Clouet, L. Pizzagalli, F. Willaime, Ab initio modeling of dislocation core properties in metals and semiconductors. *Acta Mater.* **124**, 633–659 (2017).
89. F. Maresca, D. Dragoni, G. Csányi, N. Marzari, W. A. Curtin, Screw dislocation structure and mobility in body centered cubic Fe predicted by a gaussian approximation potential. *NPJ Comput. Mater.* **4**, 69 (2018).
90. E. L. Buehler, M. J. Buehler, X-LoRA: Mixture of low-rank adapter experts, a flexible framework for large language models with applications in protein mechanics and molecular design. *APL Mach. Learn.* **2**, 026119 (2024).
91. D. B. Miracle, O. N. Senkov, A critical review of high entropy alloys and related concepts. *Acta Mater.* **122**, 448–511 (2017).
92. E. P. George, D. Raabe, R. O. Ritchie, High-entropy alloys. *Nat. Rev. Mater.* **4**, 515–534 (2019).
93. A. Ghafarollahi, F. Maresca, W. Curtin, Solute/screw dislocation interaction energy parameter for strengthening in BCC dilute to high entropy alloys. *Model. Simul. Mater. Sci. Eng.* **27**, 085011 (2019).
94. Y. Rao, C. Baruffi, A. De Luca, C. Leinenbach, W. Curtin, Theory-guided design of high-strength, high-melting point, ductile, low-density, single-phase bcc high entropy alloys. *Acta Mater.* **237**, 118132 (2022).
95. Z. Fan, B. Xing, P. Cao, Predicting path-dependent diffusion barrier spectra in vast compositional space of multi-principal element alloys via convolutional neural networks. *Acta Mater.* **237**, 118159 (2022).
96. X. Wang, L. Valdevit, P. Cao, Neural network for predicting Peierls barrier spectrum and its influence on dislocation motion. *Acta Mater.* **267**, 119696 (2024).
97. A. Ghafarollahi, M. J. Buehler, Rapid and automated alloy design with graph neural network-powered LLM-driven multi-agent systems. arXiv [Preprint] (2024). https://doi.org/10.48550/arXiv.2410.13768 (Accessed 29 October 2024).
98. R. K. Kuu *et al.*, Learning from nature to achieve material sustainability: Generative AI for rigorous bio-inspired materials design. An MIT Exploration of Generative AI, 27 March 2024. https://doi.org/10.21428/e4baedd9.33bd7449. Accessed 29 October 2024.
99. A. H. Larsen *et al.*, The atomic simulation environment–a Python library for working with atoms. *J. Condens. Matter Phys.* **29**, 273002 (2017).
100. D. Bacon, Y. N. Osetsky, D. Rodney, Dislocation-obstacle interactions at the atomic level. *Dislocations solids* **15**, 1–90 (2009).
101. E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, P. Gumbsch, Structural relaxation made simple. *Phys. Rev. Lett.* **97**, 170201 (2006).
102. H. Jónsson, G. Mills, K. W. Jacobsen, "Nudged elastic band method for finding minimum energy paths of transitions" in *Classical and Quantum Dynamics in Condensed Phase Simulations* (World Scientific, 1998), pp. 385–404.
103. G. Henkelman, B. P. Uberuaga, H. Jónsson, A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **113**, 9901–9904 (2000).
104. G. Henkelman, H. Jónsson, Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **113**, 9978–9985 (2000).
105. P. Andric, W. A. Curtin, Atomistic modeling of fracture. *Model. Simul. Mater. Sci. Eng.* **27**, 013001 (2018).
106. A. Stroh, Dislocations and cracks in anisotropic elasticity. *Philos. Mag.* **3**, 625–646 (1958).
107. Q. Wu *et al.*, Autogen: Enabling next-gen LLM applications via multi-agent conversation framework. arXiv [Preprint] (2023). https://doi.org/10.48550/arXiv.2308.08155 (Accessed 29 October 2024).
108. A. Ghafarollahi, M. J. Buehler, Automating alloy design and discovery with physics-aware multimodal multiagent AI. Github. https://github.com/lamm-mit/AtomAgents. Deposited 18 July 2024.