Review of Materials Research 1 (2025) 100010

Contents lists available at ScienceDirect

Review of Materials Research

journal homepage: www.sciencedirect.com/journal/review-of-materials-research

AI4Materials: Transforming the landscape of materials science and enigneering

Xue Jiang [a], Dezhen Xue [b], Yang Bai [a], William Yi Wang [c], Jianjun Liu [d], Mingli Yang [e], Yanjing Su [a,f,*]

[a] Beijing Advanced Innovation Center for Materials Genome Engineering, Institute for Advanced Materials and Technology, University of Science and Technology Beijing, Beijing, 100083, China
[b] State Key Laboratory for Mechanical Behavior of Materials, Xi'an Jiaotong University, Xi'an, 710049, China
[c] State Key Laboratory of Solidification Processing, Northwestern Polytechnical University, Xi'an, 710072, China
[d] State Key Laboratory of High Performance Ceramics and Superfine Microstructure, Shanghai Institute of Ceramics, Chinese Academy of Sciences, Shanghai, 200050, China
[e] Institute for Nanobiomedical Technology and Membrane Biology, Sichuan University, Chengdu, 610041, China
[f] Chinese Materials Research Society, Beijing, 100083, China

ARTICLE INFO

Keywords:
Intelligent computation
Machine learning
Materials data infrastructure
Autonomous experiment
Intelligent manufacture

ABSTRACT

New materials, crucial for economic and technological progress, are prioritized globally with strategies to accelerate their advancement through big data and AI. AI for Materials (AI4Mater) serves as an overall framework for integrating AI into Materials Science and Engineering, which is structured around three main elements: materials data infrastructure, AI4Mater techniques, and applications. This article reviews the development procedure and recent innovations in materials data infrastructure, machine learning in materials, autonomous experiment, intelligent computation, and intelligent manufacture. These efforts aim to foster open access to AI resources and enhance the collective advancement of materials science, ultimately accelerating breakthroughs and elevating the engineering application of new materials in a sustainable manner.

1. Introduction

The development of novel materials is a cornerstone of technological innovation and economic growth, driving transformative advances across diverse industrial sectors. Governments and industries worldwide are investing in strategies that integrate big data and artificial intelligence (AI) to accelerate material innovation. This has led to the emergence of AI for Materials (AI4Mater), a rapidly evolving interdisciplinary field that combines computational design, data science and experimental techniques. The integration of design, production, and application through AI, combined with its ability to enhance computational simulations and experimental workflows, is fundamentally reshaping how new materials are created and utilized [1]. The fusion of machine learning and materials computation addresses challenges in cross-scale calculations, while AI-driven automation in experimental setups increases the efficiency of material discovery. Breakthroughs in AI and robust data infrastructures will accelerate material research and development, transform industry practices, and overcome inefficiencies arross the field. This article outlines the objectives of AI4Mater, reviews recent advancements in materials data infrastructure, machine learning, intelligent computing, autonomous experimentation, and smart manufacturing. It also proposes a road map for global collaboration, including the development of unified materials data standards, integrated and shared data resources, open access to AI models and computational tools in materials science.

2. Definition and motivation

2.1. Definition of AI4Mater

AI4Mater refers to an integrated framework for applying AI into materials field. It is structured around three main elements: materials data infrastructure, AI4Mater techniques and applications domains, as illustrated in Fig. 1. At the foundation, the materials data infrastructure

* Corresponding author. Beijing Advanced Innovation Center for Materials Genome Engineering, Institute for Advanced Materials and Technology, University of Science and Technology Beijing, Beijing, 100083, China.
E-mail address: yjsu@ustb.edu.cn (Y. Su).

https://doi.org/10.1016/j.revmat.2025.100010
Received 1 April 2025; Received in revised form 5 April 2025; Accepted 5 April 2025
Available online 6 April 2025

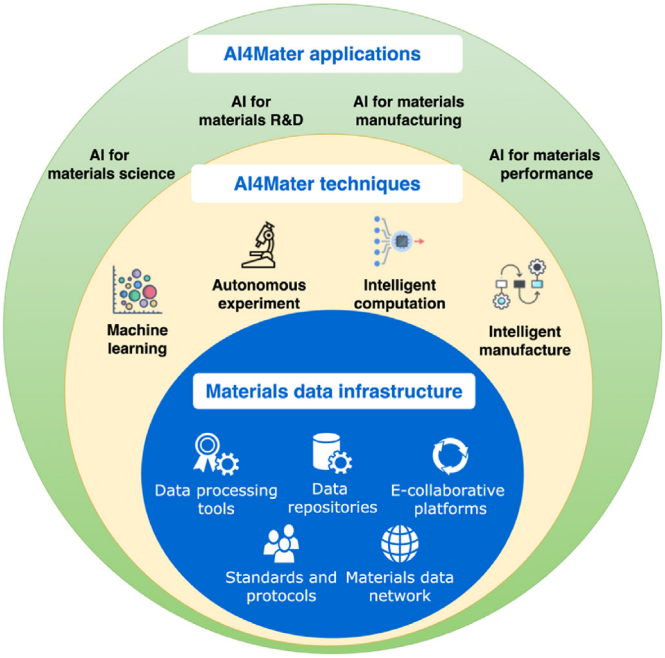

Fig. 1. The components of AI4Mater.

supports the system through tools for data processing, data repositories, e-collaborative platforms, and standards and protocols. Built on this foundation, AI4Mater integrates machine learning, autonomous experimentation, intelligent computation and intelligent manufacturing. The techniques enable applications accross material science, materials R&D, manufacturing, and performance evaluation, enhancing innovation, efficiency and scalability throughout the materials value chain.

2.2. History of AI4Mater

Over the past decade, AI has significantly advanced the landscapte of materials science, evolving from isolated apllications of isolated algorithms to a integrated, data-driven paradigm, as shown in Fig. 2. This progress can be divided into three phases:

Phase 1 (Before 2016) was characterized by an increasing awareness of new technologies, as researchers began to understand and employ various machine learning algorithms and tools. This included techniques such as regression, classification, clustering, dimensional reduction, active learning, and reinforcement learning. During this period, the focus was primarily on building foundational knowledge and developing methodologies to analyze and interpret material datasets effectively.

Phase 2 (2016–2020) was the emergence of a new data-driven paradigm in materials research. This phase was marked by the rapid adoption of data-centric approaches to material composition design, process optimization, and property prediction and improvement. Researchers began to mine structure-property relationships using large datasets, leading to more efficient materials discovery and development processes. As data generation being accelerated, the ability to leverage this information for informed decision-making became a crucial aspect of materials research.

Phase 3 (2020-Present) has been defined by an even deeper integration of AI into materials R&D. During this phase, intelligent computing, autonomous experimentation, big data analytics, and generative AI have collectively transformed the landscape of materials science. This integration represents a significant shift toward the fifth R&D paradigm in materials, emphasizing not only efficiency and optimization but also approaches to theory innovation and materials discovery. As AI technologies continue to evolve, their application in materials science promises to further revolutionize the field, enabling the rapid acceleration of material innovation and tailored solutions for various applications.

This evolution of AI in materials research highlights a continuous journey of integrating fragmented technologies into a cohesive system. The application of AI technology in materials computation, experimentation, and production has significantly enhanced the intelligence level of computational design, experimental research, and manufacturing processes. Leveraging its robust capability to handle high-dimensional data, AI effectively integrates computational, experimental, and production technologies through data-driven approaches. By establishing a closed-loop feedback iteration and autonomous optimization across the material design-experimentation-production cycle, AI dramatically improves the efficiency of new material development and engineering applications. From the initial application of basic machine learning algorithms to the sophisticated AI-driven approaches of today, each phase has built upon the previous one, fostering a more integrated and effective R&D framework. The gradual unification of AI tools and methodologies has laid the foundation for a new era of materials innovation, where AI not only supports but drives the discovery and development of new materials. This transition sets the stage for the next phase, where AI will play an even more critical role in addressing the challenges and opportunities in materials science.

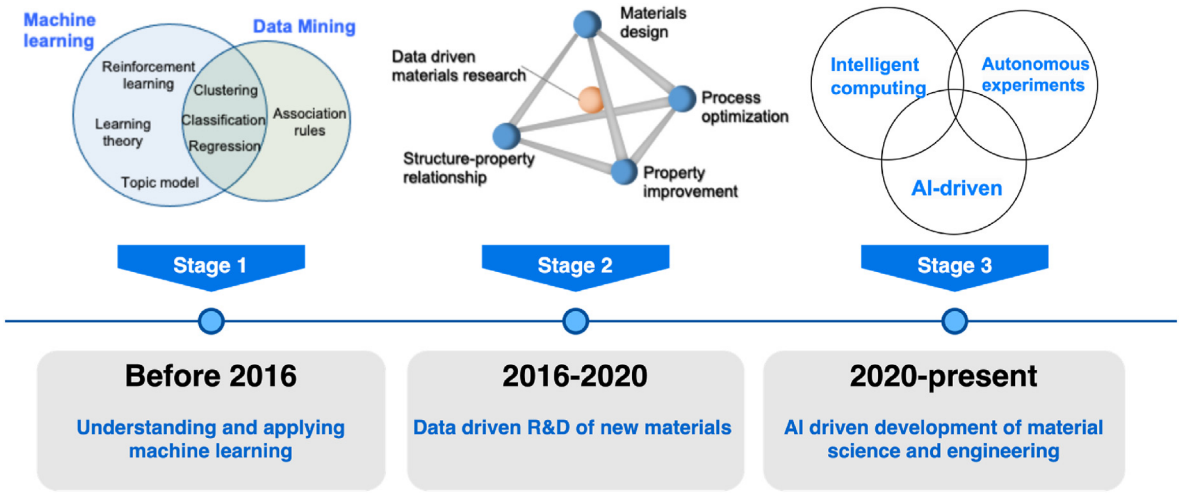

Fig. 2. The history of AI4Mater.

2.3. Motivation for AI4Mater

The discovery and development of advanced materials lie at the heart of technological progress, enabling breakthroughs in renewable energy, aerospace, high-performance electronics, and beyond. However, conventional approaches—dominated by trial-and-error experimentation and computational simulations—are slow, costly, and often constrained by existing theoretical frameworks. These limitations hinder innovation and delay critical solutions to global challenges such as energy sustainability and resource scarcity. AI4Mater is reshaping this paradigm by accelerating innovation, reducing risks, generating new scientific insights, enabling disruptive material applications, and bridging the gap between materials design and manufacturing.

Turning the Impossible into Reality: Unprecedented R&D Efficiency. AI dramatically enhances research efficiency by predicting material properties with high accuracy before experimental validation, minimizing costly trial-and-error processes. Data-driven insights enable the rapid identification of promising compositions, while AI-guided simulations and automated experimentation optimize synthesis pathways. These advancements unlock previously inaccessible material spaces, making discoveries feasible at an unprecedented pace.

Turning Hesitation into Confidence: Risk Reduction Across the R&D Chain. AI mitigates risks throughout the materials development pipeline by providing predictive insights into material performance, manufacturability, and long-term stability. Early-stage screening identifies unviable pathways before costly investments, while AI-driven process optimization enhances yield, reduces waste, and ensures sustainability. Real-time process control further enables cost-effective, scalable production with minimized economic and environmental risks.

Unraveling the Unknown: Fundamental Breakthroughs in Materials Science. AI is not just an accelerant—it is a knowledge generator. By integrating multi-scale simulations, experimental data, and theoretical models, AI uncovers hidden correlations and extracts interpretable design principles. Explainable machine learning sheds light on complex structure-property relationships, driving new scientific theories and expanding our fundamental understanding of materials.

From Discovery to Disruption: Revolutionizing High-Performance Materials. AI-powered innovation paves the way for the development of next-generation materials with unprecedented functionality—superior strength, extreme resilience, enhanced conductivity, or tailored surface properties. By accelerating the ideation-to-implementation cycle, AI enables disruptive breakthroughs in energy storage, lightweight composites, self-healing materials, and beyond.

Bridging Materials and Manufacturing: Lifecycle-Centric Material Management. AI seamlessly integrates materials design with manufacturing, ensuring end-to-end lifecycle management. From raw material selection to process optimization, from real-time monitoring to recyclability assessment, AI enables smart, adaptive control over the entire materials value chain. This convergence enhances sustainability, facilitates digital twin applications, and ensures materials innovation translates into scalable, real-world impact.

With these powerful capabilities, AI4Mater is transforming how we discover and design materials. More importantly, it bridges the gap between innovation and manufacturing, ensuring that breakthroughs in the lab translate into practical, scalable solutions. AI4Mater is not only accelerating progress—it's redefining the boundaries of what science and technology can achieve.

3. Foundations and frontiers: Global and domestic perspectives

3.1. Global perspective

The integration of AI and data-driven approaches has become a defining force in materials science worldwide, with major economies implementing national strategies to accelerate discovery, design, and application of advanced materials.

In 2011, the Materials Genome Initiative (MGI) was launched as part of the Advanced Manufacturing Partnership, aiming to halve the time and cost of new materials development through high-throughput computation, high-throughput experimentation, and materials data infrastructure. Over the years, MGI has played a pivotal role in shaping the U.S. materials innovation ecosystem, fostering collaboration among government agencies, research institutions, and industries. The initiative's Strategic Plan updates in 2014 and 2021 reinforced investments in AI-driven materials discovery, advancing intelligent materials R&D. Recognizing AI's transformative potential, the U.S. National Academy of Sciences released the 2019 Materials Research Frontiers: A Decadal Survey, highlighting AI and data science as fundamental tools for overcoming inefficiencies in materials development. More recently, the 2024 Supercharging Research report by the President's Council of Advisors on Science and Technology identified A Phase Change for Materials Discovery and AI for Designing Advanced Semiconductors as key areas where AI can accelerate scientific breakthroughs, underscoring its role in next-generation materials innovation.

Europe has long been at the forefront of data-driven materials research. In 2011, the Accelerated Metallurgy (AccMet) program, managed by the European Space Agency (ESA) and involving 31 European institutions, pioneered high-throughput alloy design and simulation, significantly reducing materials R&D timelines and costs. In 2012, the Metallurgy Europe initiative (2012–2022) identified 17 priority material needs and 50 cross-industry metallurgical research themes, covering applications in clean energy, green transportation, healthcare, and next-generation manufacturing. The European Commission's Horizon 2020 research and innovation framework (2014–2020) further integrated research efforts across EU member states, promoting cross-disciplinary collaboration and accelerating the adoption of AI-driven materials innovation. Meanwhile, Germany's Industry 4.0 strategy, introduced in 2013, emphasized the use of IoT-driven information systems to digitize and optimize production processes, laying the foundation for AI-enhanced smart manufacturing and materials design.

The UK has taken a decisive step toward AI-driven materials innovation with the launch of the National Materials Innovation Strategy in January 2025. Drafted by the Henry Royce Institute—the UK's national hub for materials science—the strategy reflects input from over 2,000 experts across 270 organizations. It prioritizes sustainability, energy efficiency, and healthcare applications, aiming to position the UK as a leader in advanced materials research. By fostering collaboration between academia, industry, and government, the strategy seeks to accelerate the translation of cutting-edge research into real-world applications. Key initiatives include the establishment of AI-powered materials innovation hubs, the development of high-throughput computational models and intelligent material databases, and the deployment of AI-optimized experimental platforms. These efforts aim to enhance the UK's global competitiveness in materials science while addressing critical challenges such as climate change and resource scarcity.

At the international level, the International Science Council (ISC) released the 2025 edition of Preparing National Research Ecosystems for AI, a report aimed at guiding nations in adapting their research infrastructure to AI's transformative potential. Published from ISC headquarters in Paris, the report provides actionable recommendations for integrating AI into national research frameworks, fostering interdisciplinary collaboration, and addressing ethical, regulatory, and infrastructural challenges. By showcasing best practices and case studies from leading nations, the ISC underscores the importance of building inclusive, equitable, and sustainable AI-driven research ecosystems to address global challenges while ensuring responsible AI deployment.

3.2. Domestic perspective

China has simultaneously developed targeted national strategies to harness AI's potential in materials science. The Chinese government has

placed significant emphasis on the development of new materials, recognizing their strategic importance for technological advancement and industrial competitiveness. In 2015, it launched the national key R&D program Key Technologies and Supporting Platforms for Materials Genome Engineering, with the ambitious goal of halving both the R&D cycle and costs — commonly referred to as the "two halves" strategy. This program focuses on establishing three major demonstration platforms to support research and innovation in materials genome engineering: high-throughput computation, high-throughput synthesis and characterization, and specialized materials databases.

To validate the effectiveness of these approaches, the program has facilitated demonstration applications across five key material categories: energy materials, biomedical materials, rare earth functional materials, catalytic materials, and special alloys. These efforts have led to significant breakthroughs, reinforcing the applicability and impact of materials genome engineering in accelerating new materials development. In 2020, further initiatives were launched to promote the adoption of materials genome engineering technologies, driving their transition from fundamental research to industrial application.

More recently, in 2024, China introduced the Construction Plan for the New Materials Big Data Center, aiming at integrating and developing material data resources to establish a robust foundation for AI-driven materials research. These initiatives have propelled advancements in fundamental materials genome engineering research, the development of key equipment and core software, and the engineering application of intelligent materials technologies. The growing synergy between material data resources and AI has played a crucial role in shaping new paradigms for intelligent materials R&D, paving the way for innovative breakthroughs.

In parallel, leading innovation hubs such as Beijing and Shanghai have unveiled their AI + Materials Action Plans, marking a pivotal step in integrating AI with materials science. These initiatives seek to accelerate the discovery, design, and application of advanced materials by leveraging AI-driven technologies. By establishing dedicated AI-powered materials innovation hubs, both cities are prioritizing the development of high-throughput computational models, intelligent material databases, and AI-empowered experimental platforms. Furthermore, these efforts emphasize cross-industry collaboration, bringing together academia, research institutions, and enterprises to build a dynamic ecosystem for AI-powered materials innovation. As these initiatives gain momentum, they are set to redefine the landscape of materials research, driving technological breakthroughs and industrial transformation in the era of intelligent materials.

4. Prior efforts on innovation

4.1. Materials data infrastructure

With the rapid growth of big data and AI technologies, the demands on materials databases have shifted significantly, evolving from simple data hosting platforms to dynamic, data-driven systems. This transformation has necessitated a new approach to database design and management, as traditional methods of predefining schemas and rigid database structures no longer meet the needs of heterogeneous and multimodal materials data. Instead, templated database systems that allow for user-defined data descriptions and flexible data integration are emerging as the new standard. This shift facilitates seamless data sharing, interconnection, and post-definition of material data schemas, ultimately driving forward the open collaboration and innovation in materials science.

4.1.1. Repositories

With the advancement of big data and AI technologies, the characteristics of material data—such as heterogeneity and multimodality—have become increasingly prominent. Consequently, the goals of materials databases are undergoing significant changes. Prior to 2010, these databases primarily offered data hosting and search services, enabling scientists to query and utilize data effectively. Additionally, they encouraged materials scientists to share their data within a broader community, including platforms like PAULING FILE, Granta Design, MatNavi, and Springer Materials. Following 2011, the U.S. Materials Genome Initiative fostered the development of a data-driven model for materials innovation, marking a transition into a new phase for materials databases. They evolved into data centers that provide raw data, computational simulations, and data analysis services, ultimately transforming into materials discovery platforms that facilitate the R&D of new materials.

This shift in demand has raised new requirements for the storage and management of materials data. The process of constructing materials databases is an abstraction process aimed at addressing materials science issues, and this process has traditionally involved two steps: (1) the abstraction from the real world to the information world and (2) the abstraction from the information world to the machine world. The first step of abstraction is carried out collaboratively by materials scientists and database engineers. Materials scientists create conceptual models that can describe the entities, properties, and relationships of materials, while database engineers formalize and impose constraints on these conceptual models to ultimately develop a logical model of the information world. The second step of the abstraction process involves database engineers using database management systems to convert the logical model into a software-supported physical model, such as MySQL, Oracle, PostgreSQL, and MongoDB. Driven by AI and machine learning technologies, the demand for the integration of materials data has intensified, leading to increasing functional requirements for heterogeneous data patterns. This makes the traditional approach of predefining materials data schemas for database construction obsolete for several reasons: (1) frequent and close communication between materials scientists and database engineers raises the design and construction barriers of material databases; (2) changes to logical and physical data storage models must be completed by database engineers, resulting in low database scalability; and (3) accessing the database through applications requires prior understanding of the database's design patterns, creating difficulties in data exchange and integration.

As a result, the field of materials databases has begun to develop templated database systems that allow for user-defined data descriptions, gradually becoming the foundation for the open interconnection and seamless sharing of multiple data resources and applications. By providing reusable data types for template customization, these systems meet the personalized data expression needs of different engineering applications, groups, institutions, and individuals, allowing for the post-definition and standardized storage of material data schemas. Examples include the Materials Data Curation System in the United States and MGEDATA in China.

The Materials Data Curation System (MDCS) [2], developed by National Institute of Standards and Technology (NIST) in the U.S., offers reusable general base types, including double-precision floating-point types, unit types, and uncertainty types. It allows users to combine these base types to create new complex data types, forming customized materials data templates for data uploads. The data is stored in a document database in the format of Extensible Markup Language (XML). MGEDATA [3,4], developed by Researchers from University of Science and Technology Beijing developed, is a materials database system. It has developed dynamic container system that supports user-defined data storage structures, providing 10 advanced general data types, including string, array, table, container, and generator types (Fig. 3). Users can combine these types to create custom data templates, and the data is stored in Binary JSON (BSON) format within a MongoDB document database, fulfilling the description and storage needs for multi-source heterogeneous materials data. Fig. 3 illustrates the graphical user interface of the dynamic container system.

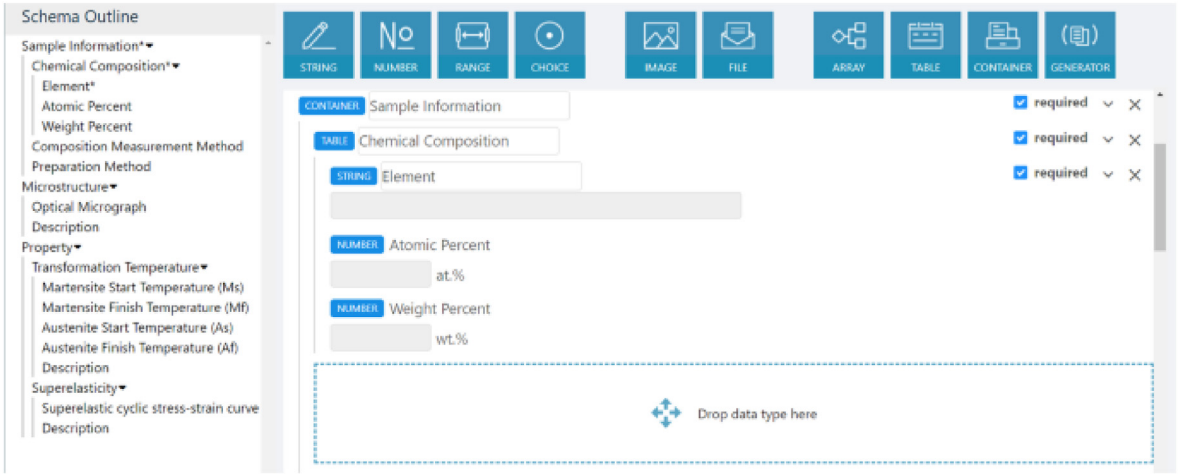

Fig. 3. The graphical user interface of the dynamic container system of MGEDATA.

4.1.2. *E-collaborative platforms*

An e-collaboration platform is a web-based environment that hosts open-source software tools, enhancing collaboration and providing distributed access to experimental and computational data. These platforms allow for the tracking of collaborative workflows, making it easier to monitor progress and contributions. Specifically designed to boost the productivity of cross-disciplinary project teams, open-source, open-access e-collaboration environments facilitate the seamless sharing and exchange of information, knowledge, and expertise among team members who are geographically and organizationally distributed across various stakeholder groups.

Since 2002, with the support of the National Science Foundation, Purdue University has established the Network for Computational Nanotechnology (NCN) and developed the nanoHUB infrastructure [5]. nanoHUB provides a self-service analysis, application development, and rapid delivery system, featuring a rapidly growing collection of simulation tools. It supports cloud-based operations and internet access, offering a user-centered, scientific end-to-end cloud computing environment. This platform serves as an open and free resource for computational research, education, and collaboration in nanotechnology, materials science, and related fields.

The U.S. government has effectively harnessed the expertise of NIST in data integration and management. Central to this initiative, NIST has partnered with nearly a hundred research institutions and companies, including Northwestern University. By consolidating materials data, codes, computational tools, and various software and hardware resources from multiple entities, they have developed a series of materials database systems, data service platforms, repositories, and resource registration systems. Notable examples include the MDCS, the Materials Data Facility (MDF), the Open Quantum Materials Database (OQMD), the NIST Materials Resource Registry (NMRR), and the CHiMaD Phase Field. Through the materials resource registration platform, they facilitate the registration of multi-source heterogeneous data, establish data exchange protocols, and bridge the gap between data resources, databases, software, and end users. This initiative paves the way for materials data integration and application both in the U.S. and globally, promoting broader sharing and reuse of materials data.

After 2011, propelled by the European Union's “Accelerating Materials” and “Metallurgy Europe” research programs, the field of computational materials data underwent rapid development. This initiative brought together computational resources and research teams from multiple European countries to establish the NOMAD database. Switzerland's National Center for Materials Computing, Design, and Discovery (MARVEL) also developed the Materials Cloud platform, offering a comprehensive solution built on open-source workflows and open datasets. This platform facilitates the sharing of computational science resources driven by materials modeling workflows, creating a new materials data ecosystem that integrates data publishing, storage, discovery, resource registration, online development, model integration, and application feedback.

In 2021, China lunched the National New Materials Database Platform (Demonstration) to explore a distributed construction concept of decentralized development and centralized sharing of materials data resources. It has created a prototype of a digital collaboration platform that integrates data accumulation, publishing and sharing, integration, retrieval and discovery, intelligent analysis, and design applications, providing cloud-based self-service analysis, tool integration, application development, and rapid delivery services for materials R&D. The design workflows provide graphical, customizable materials data processing and analysis processes, meeting the needs for low-code materials machine learning and data mining. Additionally, the platform incorporates an online coding environment using Jupyter Notebook and offers a containerized environment for materials design software and tools integration.

4.1.3. *Standards and protocols*

Since 2000, there has been an international effort to initiate and promote standardized protocols for materials data. MatML, an XML specifically developed for the storage and exchange of materials information, was initiated by NIST in 2000 [6]. MatML defines a coherent and consistent document structure for its tags, ensuring that any programming language can parse and process the data in any required manner, thereby enhancing applications' understanding and operability with the data. Initially, NIST established a working group to describe and develop the scope and specifications of the new MatML language, and released draft documents of the Annotated MatML DTD (versions 1.0, 2.0, and 3.0) for public comment. In 2003, the MatML Coordination Committee was formed, led by ASM International and the Materials Information Society, comprising experts from academia and industry. The standardization of MatML is currently managed by the Organization for the Advancement of Structured Information Standards (OASIS), with a dedicated schema development working group responsible for the development and maintenance of the MatML markup language.

MARVEL, in collaboration with 31 leading international materials data platforms, including AFLOW, the Materials Project, and NOMAD, created the Open Databases Integration for Materials Design (OPTIMADE) international alliance. This initiative introduced a universal application programming interface (API) for accessing and exchanging materials databases, promoting the FAIR principles for global materials data—ensuring that data is Findable, Accessible, Interoperable, and Reusable [7].

In 2019, the Zhongguancun Materials Testing Technology Alliance in China released the group standard "General Principles for Materials Genome Engineering Data", which standardized the information that must be collected during the data generation process. In October 2024, under the standardization system of the China Materials and Testing Standardization Committee (CSTM), the "Materials Data Standardization Field Committee (CSTM/FC93)" was established. This committee aims to develop universal technical standards, protocols, and rules covering various aspects of materials data, including storage, identification, evaluation, integration, exchange, sharing, and management.

Simultaneously, blockchain systems possess features such as immutability, complete traceability, retrievability, transparency, and collective maintenance. In the process of sharing and exchanging materials data, blockchain technology can be utilized to monitor data activities. By combining consensus mechanism algorithms and smart contract, a complete record of the data lifecycle can be established, enabling traceable and trustworthy data access, retrieval, and application. This approach has become an important data exchange model. Based on the records of data exchanges, dynamic evaluations of data resource quality and reliability can be implemented, promoting a healthy ecological cycle of data.

Materials databases and E-collaborative platforms have undergone substantial transformations to keep pace with evolving demands in the field. The integration of AI, machine learning, and big data technologies has highlighted the need for flexible, user-customizable data systems. These systems support the creation of dynamic, user-defined data templates, enabling more efficient storage, retrieval, and sharing of complex, heterogeneous materials data. E-collaborative platforms have further expanded the scope of materials R&D by offering cloud-based environments for seamless data exchange and collaboration. Meanwhile, international efforts to standardize materials data have promoted global interoperability. The combination of these advancements, along with the integration of blockchain technology for secure and traceable data sharing, is fostering a more open, collaborative, and efficient ecosystem for materials discovery and innovation.

4.2. Machine learning in materials science

The rapid advancement of AI is transforming materials science, accelerating the discovery, design, and optimization of advanced materials. Traditional trial-and-error experimentation and computational simulations often struggle with the complexity of modern materials systems. AI offers a powerful alternative by uncovering hidden patterns, predicting complex property relationships, and efficiently navigating vast design spaces. This section explores how traditional machine learning, deep learning, heuristic algorithms, generative AI, knowledge-guided models, and large language models are reshaping materials research, providing new pathways for innovation across the entire materials development lifecycle.

4.2.1. Conventional machine learning approaches for materials

The discovery and optimization of advanced materials is essential for technological innovation across industries. However, traditional process for discovering new materials has largely relied on labor-intensive trial-and-error experimentation, significantly slowing process. Machine learning (ML) has transformed this landscape by enabling faster, more accurate predictions of material properties, thereby accelerating the development process and reducing reliance on costly experimentation [8–10].

Regression models were among the earliest ML tools applied in materials science, particularly for predicting straightforward properties. Early efforts employed simple artificial neural networks to develop regression models capable of predicting properties, such as, the mechanical properties of ceramic matrix composites and the glass transition temperature of polymers [11,12]. The launch of the Materials Genome Initiative (MGI) in 2011 catalyzed a rapid expansion of ML applications in materials science, encouraging researchers to explore more advanced methods for materials prediction and discovery [7]. For example, to efficiently screen stable ternary compounds, researchers at Northwestern University employed ensemble learning to develop a model capable of predicting formation energies. This model successfully identified 4,500 new stable materials from a pool of 1.6 million candidate compositions, achieving computational speeds six orders of magnitude faster than density functional theory (DFT) calculations [13]. A landmark 2016 study published in Nature demonstrated the innovative application of support vector machines (SVMs) to guide chemical reaction design using both successful and failed experiments. This approach significantly reduced experimental workload and improved predictive accuracy, illustrating the practical benefits of machine learning in real-world materials research [14].

In recent years, machine learning-driven materials research has also flourished in China, resulting in the development of a range of high-performance advanced materials. At Shanghai University, researchers employed Random Forest (RF) and symbolic regression to predict the properties of amorphous metal alloys, successfully estimating their glass-forming ability and elastic modulus with strong predictive performance and generalizability [15]. Similarly, researchers at Northeastern University applied a general machine learning framework to predict the tensile properties of Reduced Activation Ferritic/Martensitic (RAFM) steels, demonstrating improved accuracy and broader applicability compared to traditional physical metallurgy models [16]. At Southeast University, researchers applied machine learning to screen perovskite materials, ultimately using a gradient boosting classifier to select stable candidates from 19,841 compositions. Combined with a downstream regression model, they identified 151 promising ferroelectric photovoltaic (FPV) perovskites with optimal bandgaps [17].

Today, machine learning models for materials science increasingly need to evolve into deployable tools that can seamlessly integrate into research workflows. A collaboration including the University of Chicago developed Matminer, an open-source Python software platform that provides modular tools to retrieve large datasets from external databases, perform feature extraction, and generate interactive charts for visualization. Matminer integrates closely with existing Python machine learning and data analysis packages, such as scikit-learn and pandas, streamlining data-driven materials research [18]. To further simplify the adoption of machine learning in materials science, Shanghai University created the machine learning for materials design (MLMD) platform, providing programming-free AI tools for material designers [19].

Despite the successes, challenges remain. In particular, ML models still struggle to generalize across diverse materials systems and cope with the vast, complex candidate spaces inherent to materials discovery. Future advancements will likely focus on integrating machine learning with high-throughput experimental and computational techniques, and developing adaptive models capable of learning from new data in real time while continually refining their predictions.

4.2.2. Heuristic algorithms and active learning for materials searching

The discovery of novel materials often involves navigating vast search spaces and addressing multi-objective optimization challenges. To tackle these complexities, researchers have increasingly adopted heuristic algorithms and active learning techniques. These approaches have proven particularly effective due to their global search capabilities and efficient exploration of experimental spaces.

Heuristic algorithms, such as Genetic Algorithms (GA), are rule-based methods that efficiently explore large search spaces to identify promising candidates when exhaustive searches are impractical.

Among heuristic algorithms, stand out for their robust global search capabilities. By mimicking the evolutionary processes of natural selection, GA excels at identifying optimal combinations. For instance, researchers at the University of Ottowa applied GA to optimize functional groups in metal-organic frameworks (MOFs), significantly enhancing CO_2 adsorption capability [20]. Their method screened 141

experimentally characterized MOFs and successfully identified derivatives with improved performance with CO_2 adsorption capacity increased by over 400 %. This successful application highlighted the power of GA in handling complex, high-dimensional materials optimization tasks.

In parallel, active learning, particularly within the Bayesian Optimization (BO) frame work, has become a powerful tool for materials discovery, especially in data-scarce experimental settings. BO builds predictive models that actively guide experimentation by identifying the most informative candidate materials to investigate next, thereby minimizing the number of required experiments or calculations. Researchers from Kyoto University employed support vector regression (SVR) within a BO framework, together with density function theory (DFT) calculations to optimize thermal conductivity from vast candidate pools [21]. Researchers at Xi'an Jiaotong University advanced the application of BO for material property prediction and optimization. Their work introduced an adaptive design framework to balance exploration and exploitation (as shown in Fig. 4(a) and (b)), incorporating machine learning prediction, uncertainty quantification and iterative experimental validation. It ensures efficient searches across vast material design spaces while maintaining robust performance even in data-limited conditions [22–26].

In the realm of alloy design, University of Science and Technology Beijing researchers leverage BO to optimize the composition and processing of Cu-Ni-Co-Si alloys, aiming to reduce reliance on expensive [27]. Their BO-guided optimization process required only 32 experiments to refine deformation-aging parameters, resulting in an alloy with less than half the cobalt content of ASTM C70350 while maintaining comparable mechanical performance and electrical conductivity. Building on these successes, Tsinghua University applied BO to automate materials discovery workflows, successfully optimizing CO_2 adsorption processes and screening for high-performance photosensitizers from large chemical databases [28,29]. More recently, Shanghai University and Central South University have continued to expand the applications of BO, addressing a wide range of materials design challenges [29,30].

Despite their proven advantages, both GA and BO face ongoing challenges. These include the need for large number of iterations in highly complex systems and the difficulty of managing high-dimensional, multi-component materials spaces. However, emerging approaches that integrate these methods with advanced machine learning techniques, such as deep learning-based surrogate models, are expected to further enhance their efficiency and predictive power. By providing effective strategies for exploring vast design spaces and guiding experimental campaigns, heuristic algorithms and active learning are set to remain essential tools in accelerating the discovery of next-generation materials.

4.2.3. *Deep learning for complexities in materials*

The aforementioned conventional ML methods, while effective for specific tasks, often struggle to handle the high-dimensional, complex, and large-scale data sets seen in materials science. Deep learning, inspired by the biological neural networks, has emerged as a transformative tool that can automatically extract high-level features and capture intricate nonlinear relationships through multi-layer neural network architectures [31,32]. By leveraging this ability, deep learning offers new solutions to some of the most computationally challenging problems in materials research.

A pioneering example of deep learning's impact in materials science is the Deep-learning DFT Hamiltonian (DeepH) developed by Tsinghua University. This approach directly learns from DFT (Density Functional Theory) data to predict Hamiltonian quantities (as shown in Fig. 4(c)), enabling efficient and accurate electronic structure calculations across a broad range of materials [33]. By significantly reducing the computational cost while maintaining accuracy, DeepH has been extended to applications such as magnetic superstructures and hybrid density

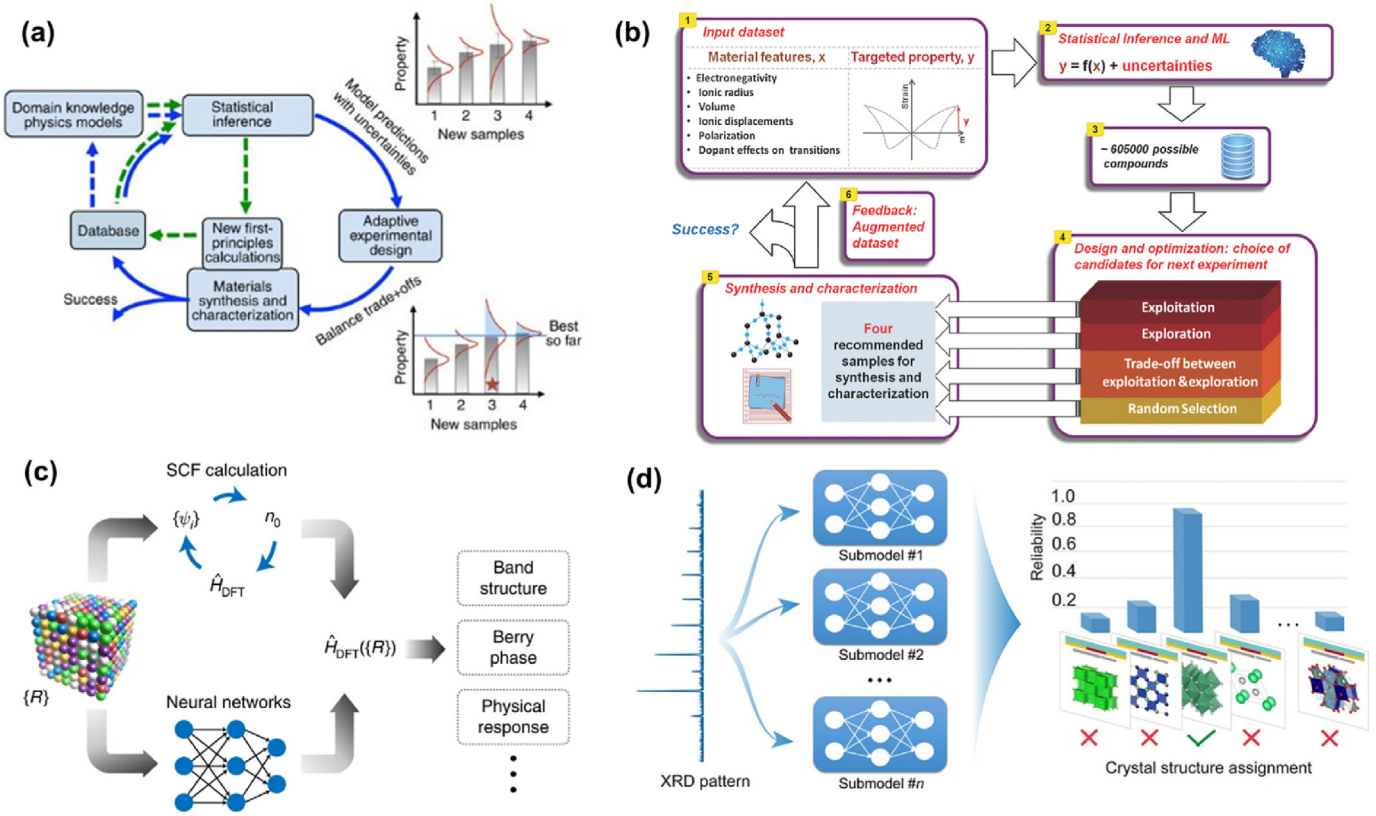

Fig. 4. (a) [22] and (b) [24] for Bayesian Optimization; (c) [33] and (d) [40] for the application of deep learning in first-principles electronic structure calculation and crystal structure assignment respectively.

functional calculations, demonstrating its adaptability and versatility in electronic structure analysis [34–36].

Deep learning has also transformed how researchers identify and analyze microscopic crystal structures. At Shenzhen Graduate School of Peking University, researchers integrated graph-theoretic mathematics with structural chemistry, creating a novel approach for structural analysis that enhances both efficiency and accuracy. This method has proven particularly valuable for tracking the structural evolution of materials during synthesis and application processes [37,38]. Further enriched by AI techniques, this graph-based methodology has been applied to lithium solid-state battery fast-ion conductors and X-ray diffraction (XRD) analysis, unlocking new opportunities for crystal

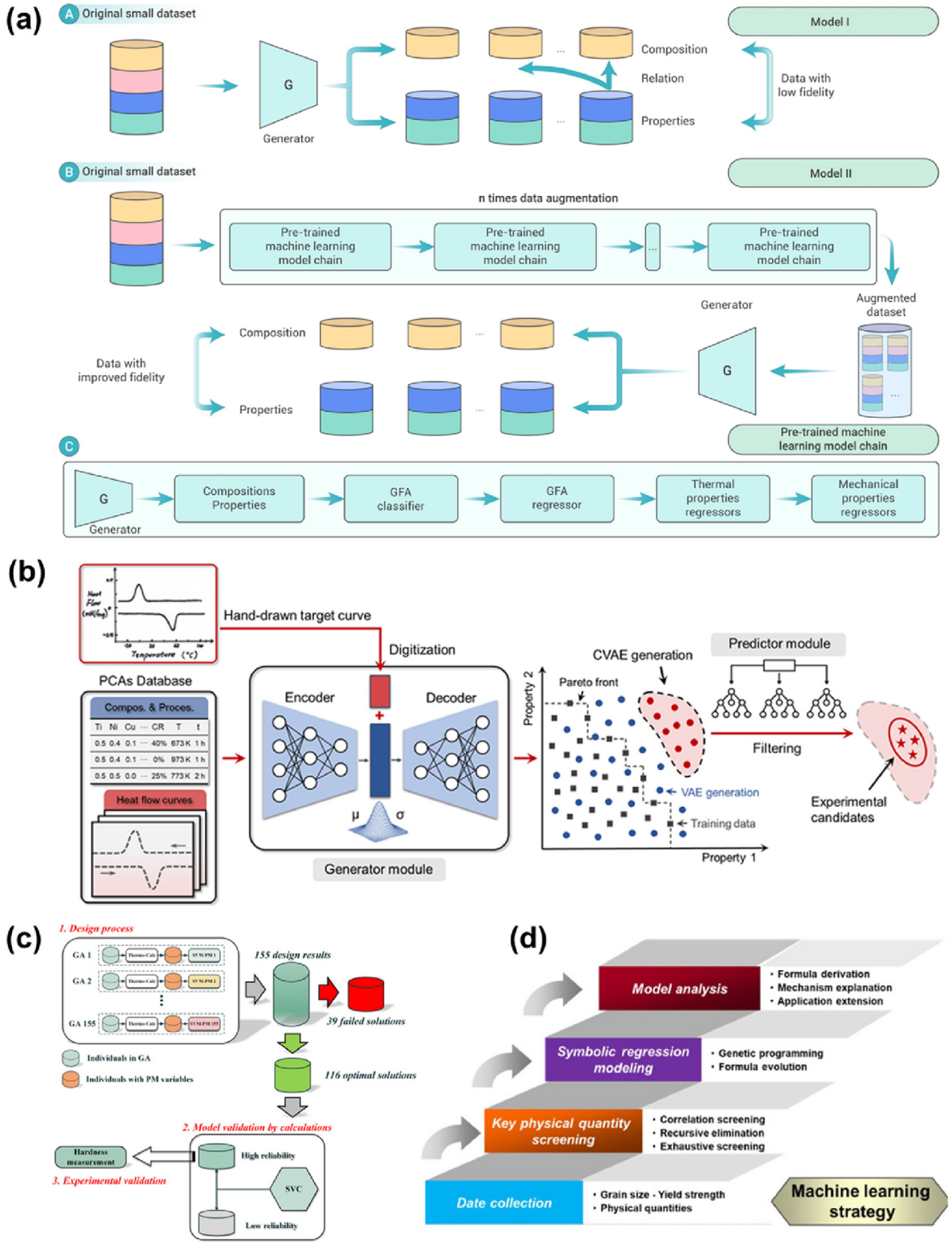

Fig. 5. (a) for GAN to copper-based metallic glasses [49]; (b) for CVAE to design SMAs [52]; (c) for physical metallurgy-guided ML approach [66]; (d) for physical mechanism knowledge integrated method [67].

structure mining and materials discovery, as shown in Fig. 4(d) [39,40].

Oriented towards property-oriented materials design, researchers at Shanghai Jiao Tong University introduced a deep transfer learning framework for predicting adsorption energies, a critical factor for understanding physical and chemical reactions at nano-interfaces. This method drastically reduces computational costs, and it has already been applied areas such as lithium-sulfur battery cathode materials and heavy metal ion adsorption processes [41,42]. In addition, by applying graph neural networks (GNNs) to model adsorption energies across various 2D materials, they identified promising catalysts for hydrogen evolution, reducing the materials screening cycle by approximately ten years [43]. This illustrates how deep learning not only enhances accuracy but also accelerates the pace of materials discovery.

Despite these transformative advances, the application of deep learning in materials science still faces challenges, particularly the reliance on high-quality, large-scale datasets, which are often scarce for novel materials. Additionally, the black-box nature of many deep learning models limits their physical interpretability, creating a gap between prediction and scientific understanding. Addressing these limitations will require the development of more efficient and interpretable deep learning frameworks, which embed materials domain knowledge into model architectures. Nevertheless, deep learning has already demonstrated exceptional potential to handle the complexity, scale, and nonlinearity inherent in materials data, and as these technologies mature, they are poised to further accelerate the discovery, design, and optimization of new materials.

4.2.4. Generative AI for inverse designing new materials

Generative AI introduced a transformative approach to materials design, enabling the inverse design of materials tailored to meet specific performance criteria. Unlike conventional ML, generative AI models usually constructs and explores a latent design space, where materials performance constraints guide the discovery of new materials. Key advances in generative models, such as Variational Autoencoders (VAE) [44], and Generative Adversarial Networks (GAN) [45], have been pivotal in optimizing material composition and designing materials structures. These foundational methods, along with techniques like Reinforcement Learning (RL) [46] and diffusion models [47], have dramatically expanded the potential for capturing complex nonlinear relationships within high-dimensional the material spaces.

Generative AI has found broad application in alloy and metallic glass design. To address the vast compositional design space in multi-principal element alloys, researchers at City University of Hong Kong applied unsupervised GAN Boosted Trees to map composition-property relationships, advancing the inverse design of bulk metallic glasses (BMGs) [48]. They later extended this approach by combining transfer deep learning with generative and supervised learning models, successfully designing copper-based metallic glasses with exceptional hardness and low modulus, as shown in Fig. 5(a) [49]. Further, the team developed an explainable AI framework integrating Conditional Variational Autoencoders (CVAE) and Artificial Neural Networks (ANN), enabling the direct generation of eutectic compositionally complex alloys across quaternary to senary systems [50]. These advances provide a new paradigm for inverse designing high-performance metallic glasses and alloys. At Xi'an Jiaotong University, researchers combined Reinforcement Learning (RL) and CVAE to efficiently explore large compositional spaces for alloy design. They applied RL to design a TiNi-based phase change alloy with the high transformation enthalpy, surpassing traditional design methods through iterative optimization coupled with experimental validation [51]. The team introduced a generative learning-enabled inverse design framework that translates hand-drawn target heat flow curves into tailored compositions and processing parameters for TiNi alloys, as shown Fig. 5(b) [52]. These works demonstrated the potential for integrating human design intent directly into machine-guided materials discovery workflows.

Generative AI has also proven valuable in designing functional materials such as high-temperature superconductors and metamaterials. Researchers at Harbin Institute of Technology, Shenzhen combined VAE and GAN techniques to systematically generate undiscovered superconductors under specific high Tc (critical temperature) constraints [48]. They later expanded this work by integrating diffusion models and a doping-adaptive representation into a three-channel matrix framework, which proposed 200 potential high-Tc superconductors not yet documented in the literature [53,54]. Their method provides a powerful tool for guiding the experimental discovery of new superconductors, opening new possibilities for material design.

In the field of metamaterials, which present even more complex design challenges due to their intricate structural requirements, generative AI has also shown promise. Researchers at Tsinghua University introduced an active learning framework that integrates VAE, multi-objective optimization, and 3D printing to design orthopedic implants with tailored microscale architectures [55]. Researchers at Huazhong University of Science and Technology developed a VAE-based generative design framework for thermal metamaterials, efficiently mapping topological unit cells to desired thermal properties [56]. These innovations illustrate how generative AI can bridge the gap between structural topology optimization and functional property design, unlocking new opportunities for advanced metamaterial engineering.

Generative AI represents a paradigm shift in materials science, offering powerful tools for the intelligent design of complex materials. As these technologies continue to evolve, they promise to further accelerate materials innovation, making the design process more efficient, accurate, and aligned with real-world applications. However, generative AI still faces challenges, particularly regarding the availability of high-quality training data and the black-box nature of many generative models, which can hinder physical interpretability. Moving forward, the integration of large language models (LLMs) for enhanced data curation and knowledge extraction, combined with expert domain input, will be critical for refining generative processes and ensuring that material designs are not only innovative but also practically viable.

4.2.5. Fusion of materials knowledge with AI

As materials science continues to embrace data-driven methods, the need to combine domain-specific knowledge with AI techniques becomes increasingly critical, especially when experimental data is scarce. While generative AI and other machine learning methods excel at navigating vast design spaces, they often lack the physical interpretability required to fully understand complex material behaviors. To address this, researchers are developing hybrid approaches that embed physical laws, expert knowledge, and theoretical models directly into AI workflows, enhancing both predictive accuracy and scientific reliability [57].

AI models are particularly effective at uncovering hidden patterns in large datasets that often remain undetected by conventional methods [58]. However, prediction accuracy can be further improved by incorporating physical features and domain-specific descriptors into these models. For example, researchers have successfully applied phase diagrams and Landau-Devonshire theory to construct new descriptors that guide the discovery of high-performance ferroelectric materials, demonstrating how physical insights can directly improve the quality of AI predictions [59,60].

To handle the vast number of potential descriptors in materials science, feature engineering techniques have been increasingly employed to select, refine, and optimize descriptors before model training. At University of Science and Technology Beijing, researchers developed a hybrid method combining feature engineering with genetic algorithms, which enabled the simultaneous optimization of machine learning model selection and material descriptors. This method achieved high accuracy in predicting phase stability in high-entropy alloys (HEAs), including both two-phase and three-phase systems [61]. They further proposed an innovative metric to quantify the efficiency of active learning (AL),

allowing the identification of optimal feature subsets, which helped accelerate the discovery of HEAs with exceptional compressive strength at elevated temperatures [62]. Another effective approach for identifying informative material descriptors comes from Shanghai University, where researchers applied the Sure Independence Screening and Sparsifying Operator (SISSO) technique [63]. Using compressed sensing principles, SISSO efficiently identifies a sparse set of descriptors most relevant to the target properties. This approach has already been applied to successfully predict perovskite stability and provided valuable insights into metal-oxide interactions, illustrating how mathematical sparsity techniques can effectively complement materials knowledge in AI pipelines [64,65].

Incorporating quantitative physical laws directly into machine learning models offers another path to improving both predictive accuracy and interpretability. Researchers at Northeastern University developed a physical metallurgy-guided machine learning model that explicitly incorporates alloy composition, crystal structure, and phase stability into its framework, significantly improving the accuracy of alloy property predictions, as shown in Fig. 5(c) [66]. Beyond prediction, AI has also contributed to the refinement of physical laws themselves. For example, the classical Hall-Petch relationship linking grain size to material strength was enhanced by incorporating intrinsic factors identified through machine learning, resulting in more accurate predictions and deeper insights into the mechanisms governing grain boundary strengthening, as shown in Fig. 5(d) [67]. In the search for novel materials, particularly in high-dimensional compositional spaces, traditional optimization techniques often get trapped in local minima. By integrating domain knowledge into AI search algorithms, researchers can constrain and guide the search process [59], steering the exploration towards physically meaningful solutions.

The fusion of domain knowledge with AI has already demonstrated success in enhancing both the predictive power and physical interpretability of machine learning models, and accelerating materials discovery. By combining physical insights, expert-crafted descriptors, and advanced algorithms, researchers can better navigate complex design spaces and uncover new materials more efficiently. As automated experimentation platforms and LLMs continue to evolve, the seamless integration of expert knowledge into AI workflows will become even more essential. Looking ahead, this hybrid approach is expected to drive new breakthroughs, enabling the rapid design of new materials tailored to meet specific performance requirements [68,69].

4.2.6. Large language models

The advancement of natural language processing (NLP) and LLMs is transforming the way we process and apply extensive materials literature, emerging as a hot topic in AI4Mater science. LLMs have been effectively utilized for the automatic extraction of materials data from journal articles, focusing on key aspects such as chemical composition, properties, and synthesis procedures. As early as 2011, researchers began to recognize chemical entities in the chemical literature, such as ChemDataExtractor [70]. Researchers at University of Science and Technology Beijing proposed an automated NLP pipeline to capture the dataset covering chemical compositions, physical properties, and synthesis and processing routes in alloys [71,72]. One notable method, ChatExtract, enables automated and precise data extraction from tables with minimal setup, utilizing conversational LLMs. By leveraging carefully crafted prompts, ChatExtract identifies, extracts, and verifies information through follow-up questions [73].

Beyond merely extracting information, the insights found in materials science literature can be efficiently encoded into information-dense word embeddings using LLMs. These embeddings can then be employed for new materials discovery and accurate property predictions. For example, researchers at Lawrence Berkeley National Laboratory have successfully adopted unsupervised learning techniques, effectively encoding scientific knowledge from materials literature into information-dense word vectors. This approach has identified relationships between material structures and properties, leading to the discovery of new thermoelectric materials [74]. Researchers at New York University also adopted the skip-gram algorithm on 6.4 million materials-related abstracts plus abstracts on metallic materials, and successfully represented high-entropy alloys by 200-dimensional word vectors. The approach identified the representative FCC Cantor and BCC Senkov alloys as the most promising high-entropy alloys, long before they had been discovered and synthesized [75].

Furthermore, LLMs accelerate materials design by automating the design, planning, and execution of complex experiments, thereby enabling autonomous research on materials. This is achieved through the integration of AI agents that utilize memory retention, logical analysis, task decomposition, and summarization capabilities. Together, these advancements highlight the significant impact of LLMs in advancing materials science and facilitating innovative research directions. Researchers at Carnegie Mellon University developed an AI system driven by GPT-4 that autonomously designs, plans and performs complex chemical synthesis experiments of organic compounds [76]. With prompt input, it can use tools to browse the internet and relevant documentation, use robotic experimentation application programming interfaces and leverage other LLMs for various tasks. Researchers from Korea Advanced Institute of Science and Technology built an AI system called ChatMOF to manage data retrieval, property prediction, and structure generations tasks by leveraging a large-scale language model. ChatMOF shows high accuracy rates of 96.9 % for text-based data searching, 95.7 % for property predicting, and 87.5 % for structure generating tasks with GPT-4 [77].

LLMs face significant challenges in directly supporting material design, particularly when it comes to quantitatively reasoning about properties based on composition and processing conditions. This raises critical questions: How accurate can LLMs be in predicting material properties, and can they actually facilitate more efficient design and optimization of materials? Researchers at Georgia Institute of Technology trained polyBERT using DeBERTa on a dataset comprising 100 million hypothetical polymers generated by exhaustively combining chemical fragments derived from a database of over 13,000 synthesized polymers. Through training, polyBERT learns to convert input PSMILES strings into numerical representations as polymer fingerprints, and a multitask machine learning framework then predicts polymer properties [78]. Researchers from the University of Science and Technology Beijing have developed an innovative end-to-end pipeline that transforms materials-related text into precise property predictions for steels, leveraging the capabilities of an LLM called SteelBERT (Fig. 6). This model has been pretrained on an extensive corpus of 4.2 million abstracts and 55,000 full-text articles in the field of materials science, specifically focusing on steel [79]. To enhance its predictive power, SteelBERT employs a disentangled attention mechanism that enables it to offer an accurate representation of steel-related knowledge when given natural language input. When examining three key mechanical properties—yield strength, tensile strength, and elongation—the researchers achieved impressive determination coefficients (R^2) of around 80 %. Furthermore, they demonstrated the model's ability to be fine-tuned with a modest laboratory dataset tailored to specific steel designs. With just 64 experimental samples of austenitic stainless steels, their pipeline effectively predicted a novel 15Cr austenitic stainless steel with yield strength of 960 MPa, tensile strength of 1138 MPa, and elongation of 32.5 %, surpassing those of all previously reported. This groundbreaking research highlights the potential of LLMs to tackle quantitative prediction challenges derived from natural language inputs, paving the way for advancements in material design across various steel types and processing routes.

The integration of AI into materials science has already delivered substantial breakthroughs, ranging from the discovery of new materials to the refinement of physical models and the development of autonomous design frameworks. As techniques such as generative AI, LLMs, and knowledge-guided learning continue to mature, AI will not only enhance

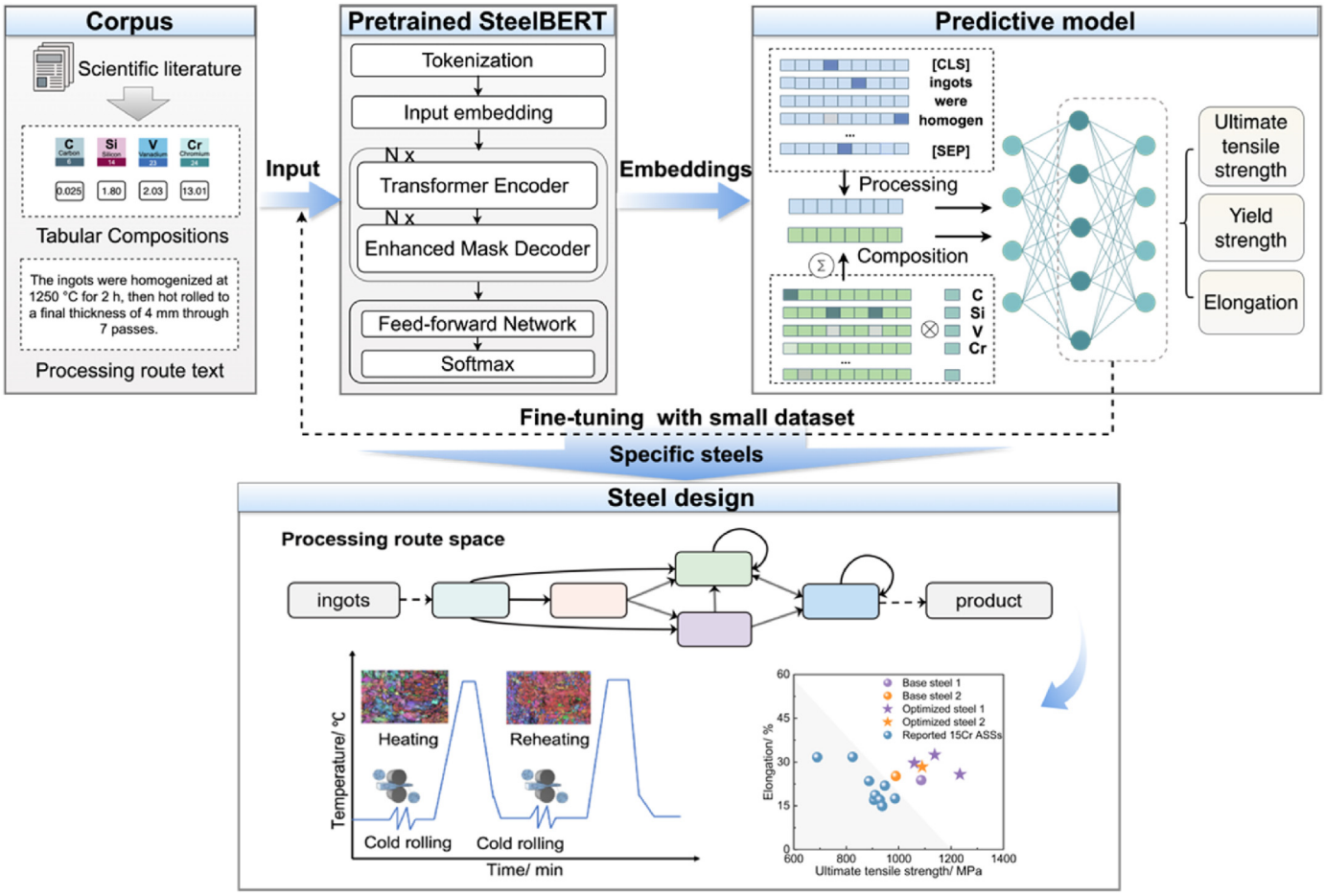

Fig. 6. Steel design based on a large language model [79].

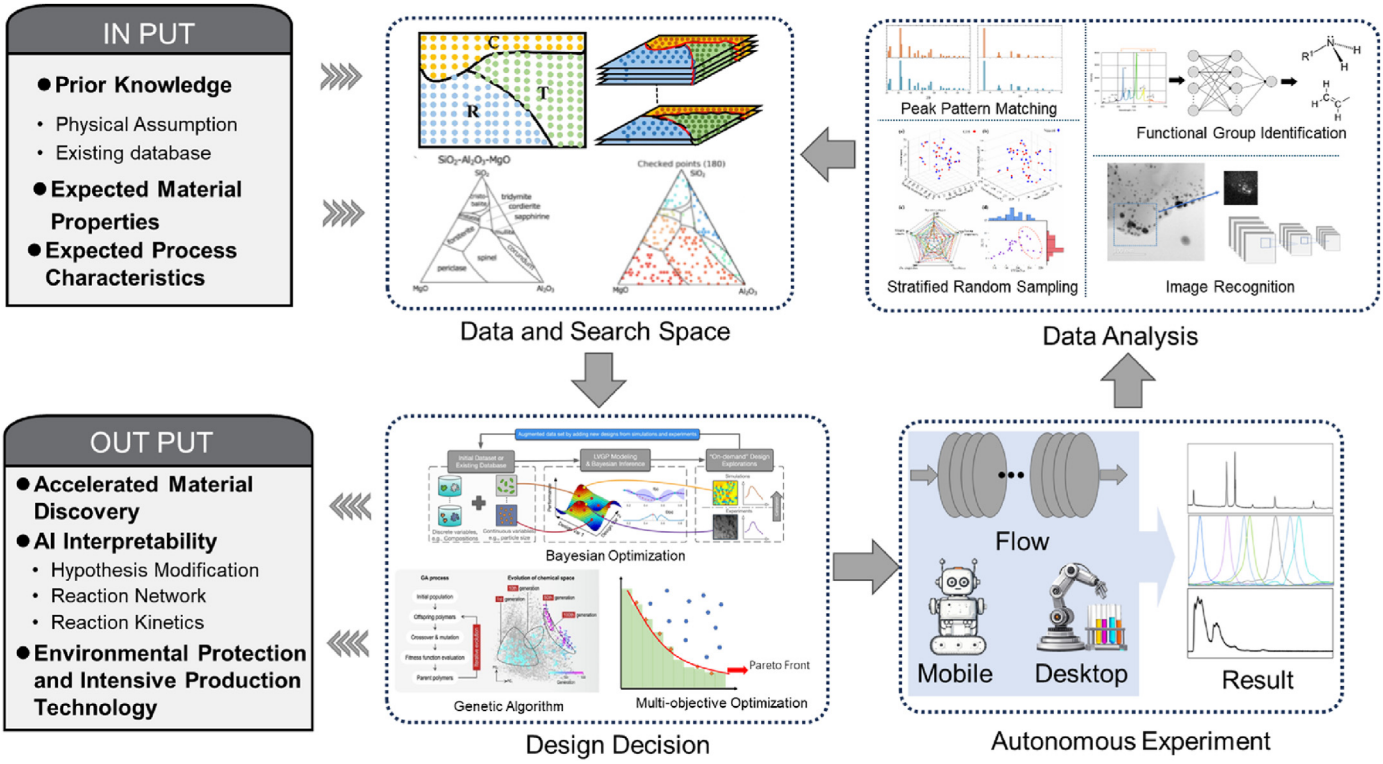

Fig. 7. The workflow of autonomous experiments [80–84].

predictive accuracy and efficiency but also deepen our understanding of complex material behaviors. Looking ahead, the seamless fusion of AI algorithms, domain knowledge, and real-time experimental data will unlock faster, smarter, and more sustainable materials innovation, ensuring that future materials are tailored to meet the ever-evolving demands of technology, energy, and industry.

4.3. Autonomous experiments

Autonomous experimentation is an innovative experimental method that integrates various advanced technologies. Its core lies in the high integration of AI, robotics, data technology, and optimization algorithms with experimental equipment, enabling the automation and intelligence of the experimental process. Fig. 7 shows the workflow of autonomous experiments. The system first uses AI to analyze and process experimental data based on predefined goals and conditions, providing a basis for the formulation of experimental plans. Then, robots perform the relevant experimental operations and collect data. The system analyzes the results again and adjusts the experimental plan. This process continues in a feedback loop, automatically adjusting the material composition and preparation process to explore new materials with better performance. Additionally, human experts can interact with the platform to provide domain knowledge and experience, guiding the optimization of experimental plans. At the same time, the system can also feedback the experimental results and analysis conclusions to the researchers, promoting mutual progress. The collaborative application of materials computation, experiment and data further enhances the intelligence level of the experimental system. On this basis, the integration of information technology, the internet, big data, and other technologies enables the modular collaborative scheduling and management of autonomous experiments, achieving efficient management and utilization of experimental and computational resources, thereby improving experimental efficiency and quality.

4.3.1. Autonomous laboratory

In 2010, the U.S. Air Force Research Laboratory collaborated with the Air Force Institute of Technology to initiate the development of the world's first autonomous materials research system (ARES). The goal was to achieve autonomous experimental design, operation, and data analysis through automated experimental apparatus and AI, significantly shortening the materials R&D cycle and fundamentally changing the paradigm of materials research. The system integrates automated experiments, AI-driven decision-making, and closed-loop data processing, achieving unprecedented research efficiency. In 2014, the results and research capabilities of ARES were announced [85]. The system is composed of a six-axis robotic arm, a multi-channel chemical vapor deposition (CVD) reactor, and an in-situ characterization module (Raman spectroscopy, SEM/TEM). By integrating hybrid AI algorithms with random forests and Bayesian optimization models, it conducted approximately 100 experiments daily, with a prediction accuracy of 92 %. The system successfully identified the temperature and hydrocarbon pressure conditions that determine the number of wall layers in carbon nanotubes from a ten-dimensional parameter network affecting their growth.

Although machine learning and predictions were performed manually offline at that time, the potential for closed-loop autonomous experimentation was already demonstrated. Using logistic regression for dimensionality reduction and Latin hypercube sampling, the system identified critical wall-layer control parameters through 237 experiments, about 300 times faster than conventional methods. This breakthrough stemmed from its innovative "parameter sensitivity stratification" strategy, which prioritizes primary effect parameters before applying response surface methodology for precision optimization. In 2016, ARES was used for in-depth studies on the growth rate of carbon nanotubes [86]. By conducting 84 preliminary experiments to gain prior knowledge, a model was established by using the random forest algorithm, and genetic algorithms were employed to determine the reaction parameters for experimental iteration. More than 600 experiments were conducted in an autonomous closed-loop mode, ultimately successfully producing carbon nanotubes at different preset growth rates. The system also reveals a completely new synthesis mechanism and discovers a new "kinetic trap" phenomenon.

In 2021, an autonomous experimental system for additive manufacturing based on ARES was released [87]. This system combined syringe extrusion printing technology with cloud-based machine learning optimization algorithms, autonomously adjusting printing parameters to achieve direct writing of single-layer printing features that meet target specifications. It employed automatic image analysis as a closed-loop feedback mechanism for online Bayesian optimization, achieving the desired printing goals in fewer than 100 experimental iterations.

Inspired by the work of ARES, some other autonomous experimental systems are designed and developed in some universities for new materials research, which can be divided into categories of flow, table and mobile according to the architecture of automation equipment. A typical example of a table system is developed by University of British Columbia in Canada, named Ada [88]. It is a modular robotic platform to accelerate the self-driving research and development of thin-film materials. This platform is driven by model-based optimization algorithms and can autonomously optimize the optical and electronic properties of thin films by modifying their composition and processing parameters. The research team demonstrated the workflow and advantages of the Ada platform by optimizing the hole mobility of spiro-OMeTAD, an organic hole transport material used in perovskite solar cells. The Ada platform, based on a modular robotic system, can automatically carry out preparation processes such as solution formulation, spin coating, thermal treatment, and property testing. The platform can install corresponding experimental modules based on specific experimental workflows, making it easily adaptable to various material research scenarios. The Ada platform features fully automated processes, allowing for the rapid generation of high-quality, well-organized datasets. It also enables the control or optimization of usually uncontrollable variables (such as the time between process steps or the height of the spin-coating dispensing nozzle), enhancing the accuracy and reliability of experiments.

Further, the hardware and software of the Ada platform was upgraded to achieve multi-objective optimization of material properties [89]. This allowed for the rapid and efficient identification of the optimal annealing temperature and conductivity for combustion-synthesized palladium thin films. By upgrading the hardware and software of the original self-driving platform, a 7-step automated workflow was implemented. A precise 4-axis robot handled the entire process of thin-film deposition and characterization, while a 6-axis robot was used to transfer samples between different modules. The preparation of four different variable thin-film samples and their performance testing were achieved through the coordination of dual-robot positions. The processing parameters and testing results were fed into the qEHVI multi-objective optimization algorithm to design the next experiment. The upgraded Ada platform allows for the control of more variables, eliminating the need for human intervention during the experimental process. It can run 40–60 experiments autonomously, reducing labor costs, minimizing human error, and ensuring the accuracy and stability of the experiments.

A typical example of a mobile system is the “Mobile Chemist” for photocatalysis experiments developed by the University of Liverpool in the UK [90]. This mobile robot can move freely within the laboratory and utilizes a combination of laser scanning and touch feedback to achieve precise positioning within a laboratory space measuring 7.3 m $\times$ 11 m, with spatial positioning accuracy of $\pm$0.12 mm and orientation positioning accuracy of $\theta \pm 0.005^{\circ}$. With high-precision control, the robot can perform dexterous operations at various stations in the laboratory, achieving a level of execution precision comparable to that of researchers. The robotic system can simultaneously consider ten dimensions of variables and operates for 21.5 h a day (with the remaining time used for charging). It significantly enhances experimental efficiency, being 1000 times faster than manual methods and at least 10

times faster than semi-automated robotic workflows. this system integrated with Bayesian optimization independently completed 688 experiments and identified the optimal catalyst formulation within an 8-day experimental cycle, increasing the hydrogen production rate by six times. Additionally, the platform can explore a larger experimental space and handle complex multi-component experiments. It is highly flexible, capable of operating in dark environments suitable for photosensitive reactions, and meets safety standards, allowing it to work in the same space as humans. Its modular design makes it easy to expand with other experimental modules.

Collaborated with the University of Liverpool in the UK, East China University of Science and Technology and Zhejiang Normal University established a materials acceleration platform for discovering molecular nanocluster photocatalysts [91]. This platform combines combinatorial molecular libraries, high-throughput automated screening and flow synthesis technology, accelerating the discovery and preparation of molecular nanocluster photocatalysts for hydrogen evolution. The advantage of this platform lies in effectively bridging the gap between small-scale high-throughput screening and large-scale production. The optimized molecular nanocluster MTPA-CA:CNP147 exhibits excellent hydrogen evolution performance. Theoretical calculations revealed the relationship between the moderate binding energy between the molecular donor and acceptor and the high catalytic activity. This research opens up a new paradigm for integrated research and production, and provides an important technical pathway for the large-scale preparation of efficient catalysts in the future.

Similarly, the University of Science and Technology of China developed an AI Chemist system and created a data-driven, automated chemistry platform [92]. This platform integrates literature reading, autonomous experimental design, chemical experiments, and chemical product optimization, aiming at studying catalysts for the automatic synthesis of oxygen-evolving reactions from Martian meteorites. The system consists of mobile robots, a computational "brain", cloud servers, and 14 dedicated workstations. It achieves full automation from ore analysis to catalyst optimization through a dual-layer workflow. The system efficiently screens formulations and quickly identifies the best catalyst from over 3 million possible compositions, significantly improving the efficiency compared to traditional trial-and-error methods. The screened catalyst can stably and efficiently produce oxygen under simulated Martian conditions, providing a reliable solution for oxygen supply in Mars exploration. Moreover, the system is highly automated, with the entire experimental process eliminating no human intervention, reducing human error and improving experimental accuracy and efficiency. It is also highly versatile, and the established experimental processes and systems can be applied to other extraterrestrial material research and chemical synthesis, advancing the development of extraterrestrial resource utilization.

A representative of flow system is "Artificial Chemist" developed by North Carolina State University, which adopted a smart-flow-based quantum dot synthesis strategy [93]. This system consists of precursor preparation, flow reaction, and in-situ characterization modules. The precursor preparation module includes multiple components that can precisely mix reaction reagents; the flow reaction module uses a flow reactor to carry out the reactions; and the in-situ characterization module uses a specially designed short-path flow cell to perform precise in-situ measurements of high-concentration quantum dots without disrupting the flow uniformity, solving the problem of significantly diluting the reaction phase for previous in-situ photoluminescence quantum yield (PLQY) measurements of quantum dots. The system can quickly explore the chemical space of colloidal quantum dots, learn synthesis pathways, identify the components and synthesis routes required to achieve specific optoelectronic properties, and archive and transfer the knowledge generated internally to subsequent experiments. It can continuously synthesize rapidly optimized quantum dots as needed, producing various customized quantum dots using a small amount of starting solution in a short time. Additionally, it can autonomously discover the best synthesis path and, through the same modular flow-synthesis platform, continuously feed reactants to achieve the large-scale production of target perovskite quantum dots.

In 2024, the University of Amsterdam in the Netherlands designed a multifunctional RoboChem robotic platform based on a flow system for the self-optimization and process intensification of photocatalytic reactions [94]. Users input experimental parameters through a graphical user interface (GUI), and the experimental equipment, controlled by Python code, begins to operate. The obtained experimental data is fed back to a Bayesian optimization algorithm in real time. This reduces manual intervention, decreases dependence on specialists, and enhances experimental safety. At the same time, the platform efficiently explores vast reaction parameter space, accurately identifies the optimal reaction conditions for different substrates, significantly improves the reaction yield and space-time yield. Additionally, its modular design facilitates future integration with other equipment, and the accumulated high-quality experimental data contributes to the advancement of the digitalization of synthetic chemistry.

Combining with 3D printing technology is another important development direction of autonomous experimental system. Boston University in the United States developed a Bayesian autonomous experimental researcher (BEAR) for the design of mechanical metamaterials in the vastness of additive manufacturing design space [95]. This system integrates five fused deposition modeling (FDM) 3D printers within the working radius of a six-axis robotic arm and is equipped with a balance to weigh and a universal testing machine to test toughness. Custom software controls the coordinated operation of these devices. This system enables the autonomous optimization of metamaterials and their high-throughput experiments, resulting in a rapid iteration to find optimal mechanical performance. It quickly quantifies the uncertainty inherent to properties such as toughness and explores the correlation between structure and performance. Therefore, an almost 60-fold reduction was achieved in the number of experiments needed to identify high-performing structures relative to a grid-based search. It provides new pathways for research in the mechanical fields and can help discover new properties when combined with methods like topology optimization.

4.3.2. *Software platform*

In 2018, the ARES development team pointed out that autonomous experimental systems could lead to discoveries of new materials growing at a rate akin to “Moore's Law”, enabling them to complete research tasks in multidimensional parameter spaces faster than human researchers and more effectively address the increasingly complex and high-dimensional demands of materials development [87]. In 2021, the open-source operating system ARES OS, capable of fully controlling and driving ARES, was introduced, featuring three main modules: interfaces, data analysis, and design planning. ARES OS consists of three core modules [87]. The interface module enables seamless communication between ARES OS and external devices and software systems, ensuring accurate data transmission and precise command execution. The data analysis module uses advanced algorithms to process experimental data, identify patterns, trends, and potential issues, and provide insights for experimental optimization. The design planning module formulates experimental strategies by considering experimental goals, available resources, and data analysis results, such as designing the best reaction conditions in carbon nanotube synthesis. Currently, ARES OS has been utilized in autonomous research systems for carbon nanotube synthesis, flow chemistry, and additive manufacturing.

In 2018, International Business Machines Corporation (IBM) created the RXN online chemistry platform [96]. Through independent analysis of literature, the platform autonomously learns reaction rules, predicts both forward and reverse reactions along with their yields for iterative cycles, formulates the synthesis pathway for the target compound, and automates the synthesis process. A key advantage of RXN lies in its Transformer model, which is grounded in the Simplified Molecular Input Line Entry System (SMILES) database. This model can be trained in an

unsupervised manner using an attention mechanism to rapidly and accurately establish atomic mappings to achieve desired outcomes. Additionally, the platform is initially equipped with cloud AI-driven and remote-control functionalities.

4.3.3. Large language models driven autonomous experiments

Recently, the rise of large language models has brought new development directions and strong help to autonomous experiments, In 2023, University of California introduced the A-Lab, an autonomous laboratory for the solid-state synthesis of inorganic powders [97]. This platform uses computations, historical data from the literature, machine learning and active learning to plan and interpret the outcomes of experiments performed using robotics. Over 17 days of continuous operation, the A-Lab realized 41 novel compounds from a set of 58 targets including a variety of oxides and phosphates that were identified using large-scale ab initio phase-stability data from the Materials Project and Google DeepMind. Synthesis recipes were proposed by natural-language models trained on the literature and optimized using an active-learning approach grounded in thermodynamics. Analysis of the failed syntheses provides direct and actionable suggestions to improve current techniques for materials screening and synthesis design. The high success rate demonstrates the effectiveness of artificial-intelligence-driven platforms for autonomous materials discovery and motivates further integration of computations, historical knowledge and robotics.

Carnegie Mellon University developed an AI system named Coscientist, leveraging advanced large-scale language models such as GPT-4 [76]. This system is capable of autonomously completing a whole set of processes, from retrieving information, planning and designing experiments, writing programs, remotely operating automated systems to do experiments, and then analyzing data. The software modules of Coscientist can retrieve publicly available information about compounds from the Internet, document data, and other available sources. After learning, it formulates synthesis pathway, determines the experimental protocol, and remotely operates the pipetting robot to conduct the experiments. Furthermore, this system utilizes publicly available chemical information in the Simplified Molecular Linear Input Specification (SMILES) format to enable communication between different modules by writing code autonomously. Coscientist combines laboratory automation with large language models, breaking through the traditional experimental design mode and opening a new direction for the intelligence of chemical experiments.

4.4. Intelligent computation

Materials intelligent computation and design refer to the application of AI in multi-scale computing, high-throughput computing, and integrated computational materials engineering (ICME). By utilizing machine intelligence, it achieves autonomous setting of trial-and-error spaces, selection of computational approaches, adjustment of computational resources, and optimization of screening criteria during the computations for new material design, preparation, processing, and deployment. This forms a closed-loop of autonomous virtual iterations to obtain optimal designs that guide experimental research and engineering applications of new materials. Intelligent computational technology breaks through the fixed thinking patterns of the human brain, autonomously explores new variables beyond known variable spaces, and autonomously optimizes design ideas and technical pathways, significantly improving material computational efficiency and enhances the reliability and accuracy of material design and screening.

4.4.1. Machine learning interatomic potentials

Traditional molecular simulations rely on quantum mechanical calculations and potential energy functions obtained by fitting experimental data. The computational cost and efficiency become a bottleneck as the scale and complexity of material systems continue to increase. The rise of machine learning interatomic potentials (MLIPs) has provided new ideas for addressing these issues [98,99]. MLIPs utilize ML algorithms to automatically learn potential energy functions from data, enabling rapid and accurate prediction of interactions between particles. Compared with traditional quantum mechanical methods, MLIPs not only offer significant advantages in computational efficiency but also adapt to a wide variety of complex materials, effectively capturing intricate interactions that traditional potentials struggle to describe.

The initial machine learning potentials primarily utilized ANNs to handle specific systems, directly using all atomic coordinates for potential energy predictions [100]. Although these models found application in certain fields, they were limited to specified systems from which the models were derived. Consequently, their universality and accuracy has restricted their applications in wider fields. To address these challenges, deep learning algorithms have emerged and gradually taken a dominant role in MLIP models. Deep Neural Networks (DNNs) algorithms can capture the complex interactions between atoms through the nonlinear mappings of multiple layers of neurons, showing exceptional performance, especially when handling large-scale datasets and high-dimensional systems. Furthermore, graph neural networks (GNNs), as an emerging deep learning algorithm, utilize graph-structured data to automatically identify interatomic relationships, further improving the MLIP accuracy.

With the development of deep learning algorithms, a variety of deep learning-based MLIP models have been developed. Among them, DP Technology® in China has developed DeePMD-kit [101], a tool using DNNs to learn the complex patterns of many-body interactions and accurately predicting the potential energy of material systems, which surpasses the limitations of many traditional classical potential models. The “Deep Potential” team, composed of Chinese and American scientists, combined molecular modeling, machine learning, and high-performance computing to extend the limit of molecular dynamics simulations to the scale of 100 million atoms [102], reducing a computation task that would normally take 60 years to just one day. This achievement earned them the 2020 Gordon Bell Prize [103]. Combined with high-performance computing, the team recently extended the ability of first-principles accuracy molecular dynamics simulations to the systems containing billions of atoms [104]. GPUMD [105] is a software package specifically designed for large-scale atomic simulations, using a single-layer neural network to construct precise MLIPs. By combining GPU acceleration with the neural network potentials (NEP, this method enables efficient prediction of atomic interactions when handling complex systems. Despite the relative simplicity of the model, GPUMD is able to provide accurate simulation results for various materials and their applications. As an emerging field, MLIPs have demonstrated great potential in the field of material simulations. With the continuous advancement of ML algorithms, datasets, and computational resources, MLIPs will play an increasingly important role in the future materials R&D.

4.4.2. Efficient computation and screening

Traditional computational methods face significant challenges when dealing with complex data and high-dimensional problems, particularly in terms of computational efficiency and accuracy. The integration of AI with computational methods, especially in areas such as high throughput computation (HTC), molecular dynamics simulations, and DFT, is driving innovation in material computation and screening.

The development of AI, especially ML models, has provided new momentum for DFT, significantly improving the computational efficiency of DFT, enabling ML models to predict the properties of new materials several orders of magnitude faster than traditional DFT methods [34,35,106]. This accelerated capability for material discovery and screening is attributed to the ability of ML models to learn from DFT-generated data, which in turn supports the rapid iteration of experimental designs and the development of automated experimental processes. In terms of accuracy, ML models, through training on large datasets, reduce prediction errors, thereby enhancing the precision of

DFT predictions. As the training dataset grows, the accuracy of the model approaches the noise level, particularly in compound screening and ranking, where its importance becomes evident. By optimizing feature sets and incorporating physical principles, ML models effectively mitigate the impact of data imperfections, thereby improving the reliability of DFT predictions in material applications. AI also enhances the transferability of DFT by learning electronic structure information. For example, researchers at Tsinghua University combined deep learning with DFT to train a neural network that efficiently represents the DFT Hamiltonian, bypassing the computationally intensive self-consistent field (SCF) iteration process [33]. This approach demonstrated high accuracy, good transferability, and high efficiency across different material systems, particularly in the study of twisted van der Waals materials. In addition, AI techniques improve the computational capability of DFT when applied to large-scale systems by breaking down tasks and utilizing local contributions. ML models can learn from localized atomic environments and extend this knowledge to larger systems, enabling in-depth studies of complex material systems.

As the complexity of molecular systems increases, traditional MD simulations face significant challenges, including a huge demand for resources and difficulty in data processing. In recent years, AI technologies have offered new approaches to address these challenges [107, 108]. First, AI enables the efficient processing and analysis of large datasets generated by MD simulations, which often contain underlying patterns and relationships that traditional methods are unable to effectively extract. Through automated data processing, AI trains models to recognize these latent relationships, thereby significantly enhancing research efficiency. For instance, researchers at Soochow University employed a fully connected neural network model to analyze the relationship between the structure and conductivity of 180,000 alkaline anion exchange membrane variants, quickly identifying promising membrane materials [109]. These predictions provided more precise initial conditions for MD simulations, effectively avoiding the cumbersome assumptions and trial-and-error processes inherent in traditional methods. Second, AI plays a critical role in optimizing parameter selection in MD simulations. In traditional MD simulations, the accuracy and computational efficiency typically depend on the precise specification of parameters such as time step, temperature, and force field. However, the selection of these parameters often requires extensive trial and error and empirical knowledge. By training ML models on historical data, AI learns from past simulations and predict optimal parameter settings based on simulation goals and the characteristics of the studied system. The integration of AI with MD simulations enhances both the efficiency and accuracy of traditional MD methods, enabling researchers to more effectively handle complex material systems.

HTC involves large datasets and complex computational processes, particularly when dealing with high-dimensional data, which often leads to the 'dimension disaster'. ML offers powerful support in this context. By adopting a data-driven approach, ML can extract valuable patterns and insights from vast amounts of data, thereby further optimizing the HTC workflows and enhancing the efficiency and accuracy of material screening and design [110]. For example, ML analyzes historical data to identify key variables, optimize the selection workflows. Moreover, AI is capable of recognizing previously overlooked potential factors, thereby refining the scientific basis of materials design. For material screening, ML models are trained on existing datasets to evaluate and rank candidate materials, enabling the identification of promising candidates within a significantly smaller pool. This new technique may provide precise assessments during the early stages of material design, eliminating the need for labor-intensive manual selection. Researchers from DeepMind [111] utilized a GNoME model, which was rained on a large and diverse set of first-principles calculations, to predict 2.2 million crystal structures, of which 380,000 have been confirmed as stable. This model increased the number of stable crystals by more than an order of magnitude, opening a new approach to material discovery.

One significant achievement of Materials Genome Engineering is the platform of ALKEMIE [112], acronyms for Artificial Learning and Knowledge Enhanced Materials Informatics Engineering, which facilitates the widespread of data-driven techniques across diverse scientific communities based on the design principles of AMDIV (automation, modular, database, intelligence, visualization). All structured and multi-modal data from computational processes and results are stored in the ALKEMIE-Matter Cloud, where they can be accessed, queried, and shared via high-level APIs. The Matter Cloud database currently contains over 600,000 crystal structure entries, more than 10,000 phonon spectrum records, over 100,000 DFT band structure datasets, more than 1,000 GW-accuracy band structure datasets, and over 200,000 sets of machines learning potential data. The MatCloud platform [113], developed by researchers from Chinese Academy of Sciences, employs quantum mechanics/machine learning (QM/ML) methods to predict material properties and drive the discovery of new materials, providing a fully automated solution for the entire process from modeling to data storage.

Under extreme conditions, the Professional Materials at Extreme (ProME) platform has been setup, which is consisted of several toolkit with IPR (Independent Intellectual Property), as shown in Fig. 8. It can be seen that Similar Atomic Environment (SAE) tool and AI-based Crystal (ABC) search tool are designed to construct the order/disorder crystal structures, whose atomic and electronic structures together with some physical properties can be precisely predicted via the classical first-principles calculations [112]. Through utilizing the further analysis tools such as the second version Mean-Field Potential (MFP^2) tool and Auto-Calphad, thermodynamic properties especially for those extreme conditions can be obtained. The High-performance Tool of Elastic Modeling (HTEM) and Efficient Calculation of alloy by Coherent Potential Approximation (ECPA) tools display their capability in predicting the elastic and plastic properties while HiPPOS (High strain rate Plastic Phase field mOdel) together with integrated models and algorithms simulate the microstructure evolutions and optimizations. The Corrected Assessments & Self-adaptive Tool (CAST) is designed as the fundamental one organizing the workflows and smartly corrected the errors during high-throughput first-principles calculations. It is worth mentioning that all these individual tools together with the integrated ProME platform have been tested by third party and validated and verified in various materials.

4.4.3. *ICME in AI $^+$ era*

ICME integrates computational modeling, experimental validation, and engineering design, employing multi-scale modeling techniques to unify physical processes from the atomic level to the macroscopic level. This approach enables researchers to optimize processing techniques and predict product performance prior to materials fabrication and product manufacturing. ICME covers key components of product development, including materials property databases, microstructural models, microstructure–property models, and cost analysis models [115]. However, as materials design becomes more complex, traditional ICME approaches face significant challenges when dealing with high-dimensional, multivariable, and nonlinear behaviors. The integration of AI offers new solutions for ICME [114]. By learning from large-scale experimental data and simulation results, AI can automatically identify underlying patterns, optimize model parameters, and thus improve predictive accuracy and computational efficiency. AI not only compensates for the limitations of traditional physical models but also handles the interaction of multiple variables and complex relationships, positioning itself as a pivotal technology driving the further advancement of ICME.

In recent years, the integration of AI and ICME has brought revolutionary changes to materials design. AI helps ICME with revealing the relationship between microstructure and macroscopic property, analyzing and learn from large volumes of experimental data and simulation results, and enabling more accurate property predictions [116]. AI has demonstrated unique advantages, particularly in situations where explicit physical models are lacking. In multi-scale modeling, AI-aided ICME integrates cross-scale data and automatically adjust model

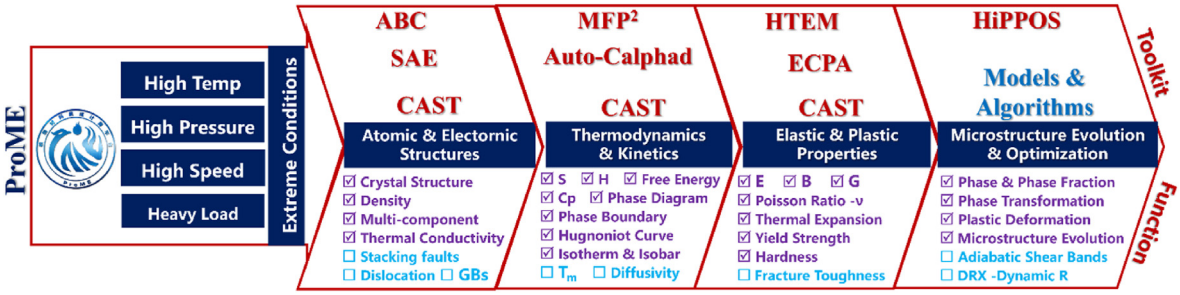

Fig. 8. Tool kits together with their functions embedded in the ProME platform with IPR (Independent Intellectual Property) [114].

parameters, enhancing the accuracy and adaptability of the models [117]. In materials design and optimization, AI helps quickly identify promising material combinations from vast amounts of experimental and simulation data, and automatically optimize microstructure, composition, and processing conditions [118]. This capability is especially valuable in the development of materials for high-performance or specific applications.

The integration of AI and ICME has driven a leap forward in materials science. For example, researchers at Central South University [119] used a multiple loops machine learning framework to establish the composition–processing–structure–property relationship of cemented carbides, enabling the intelligent design. Researchers at Northwestern Polytechnical University [120] adopted a data-driven ICME framework to achieve intelligent design and optimization of advanced structural metal materials, such as light-weight materials (e.g., Ti, Mg, and Al alloys), refractory high-entropy alloys, and superalloys. These successful applications highlight the significance of knowledge-based modeling/simulations, machine learning algorithms, and the ICME ecosystems. Researchers at Texas A&M University [121] have introduced a method that integrates phase-field modeling with deep Variational AuToEncoder for inverse microstructure-centered materials design. This method has been successfully applied to predict the effective thermal conductivity of multi-phase materials.

Moreover, it is noted that AI aided design and manufacturing have been considered as one kind of robust data-driven and data-intensive technologies in the ICME era [115,120]. As presented in Fig. 9, the evolution of smart design and manufacturing paradigm of advanced materials from knowledge-enabled data-driven ICME [115,120] to AI$^+$ era, promoting the transformations from designing the materials to designing with materials and driving the dramatically evolution of AI$^+$ paradigm from the Embedding AI to the Copilot AI and the on-going AI Agent [114]. While the MGI highlights the dominate roles of experimental tools, computational tools and databases together with their interactions, the HCPs & MGE emphasis on the interactions of Human-Cyber-Physics [122], the common cross field of which indicate the future applications of those advanced AI technologies. The machine learning algorithms could accelerate the generation of fundamental materials insights and basic science research through identification of main/value data relationships to strengthen human-physics and human-cyber interpretations and yield scientific knowledge/model [123]. Through digitization of knowledge, the Cyber-Physical systems (CPs) can be integrated by developing tools to digitize the knowledge and to crat new materials together with their robust manufacturing process [124]. Industry 4.0 fueled by data and machine learning is building on the digital advances by bridging the physical and digital world through cyber-physical systems, which has transformed into Materials 4.0 [124]. Comparing the top-down engineering routine with the bottom-up computational design one, the twin features between experimental and theoretical chains are highlighted by two types of arrows with different background colors, which also reveal the corresponding composition-processing-structure-properties-performance (CPSPP) relationship during the smart design and manufacturing. On one hand,

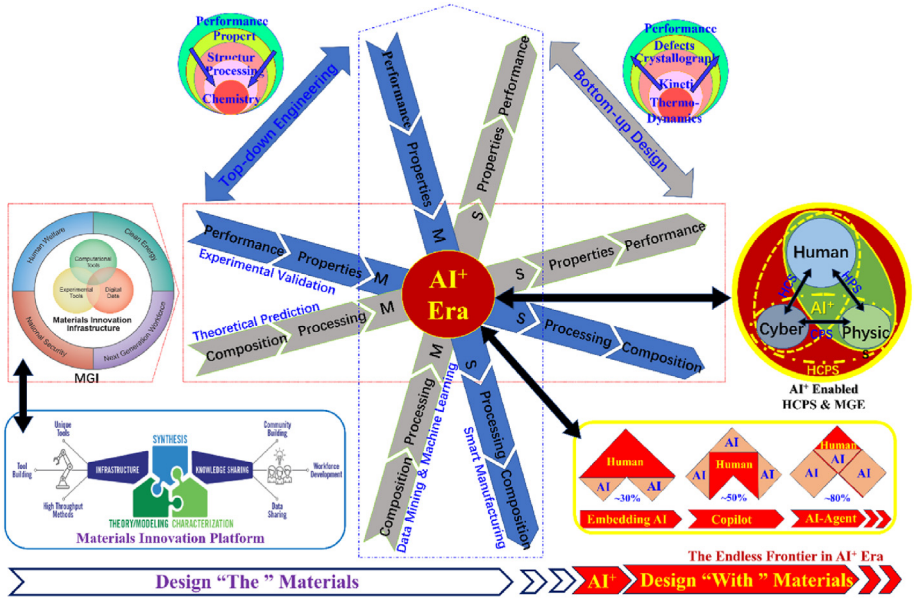

Fig. 9. The evolution of smart design and manufacturing paradigm of advanced materials from knowledge-enabled data-driven ICME [115,120] to AI$^+$ era, promoting the transformations from designing the materials to designing with materials and driving the dramatically evolution of AI$^+$ paradigm from the Embedding AI to the Copilot AI and the on-going AI Agent [114]. The insert description of Materials Innovation Platform was reported in Ref. [132].

digital twins have been considered important tools for realizing the real-time interaction and integration between information and the physical world, and represent a key enabling technology for achieving cyber-physical integration for smart assembly/manufacturing [125], i.e., for aircraft, trains and engines [114,120,126]. On the other hand, according to the CPSPP relationship, process-structure and structure-property models highlight the significant role of materials knowledge in the transformation from data and informatics to manufacturing, which also indicate the high-value chain toward intelligent manufacturing [125,126]. Furthermore, motivated by the LLM and the generative and the next-generation AI [14,127–129], the role of AI in the assignments of accelerating materials discovery and manufacturing would be increased from ~30 % of Embedding AI paradigm to ~50 % of Copilot AI paradigm, and even ~80 % of on-going AI Agent paradigm [114]. Therefore, through accelerating materials discovery and manufacturing enabled by LLM or AI-agents [130], the Human-Cyber interactions in line with the AI-enabled HCPs and MGE will be dramatically changed. The materials S&T and the scientific discoveries and applications in AI^{+} era will transform from "design the materials" to "design with materials" [114,130,131].

4.5. Intelligent manufacture

4.5.1. AI-assisted 3D printing

The traditional subtractive manufacturing process often results in materials waste and has limitations in terms of precision and cost control when processing complex shapes. 3D printing builds three-dimensional objects layer by layer, directly fabricating complex geometries from digital design files. This technology surpasses the size and shape constraints of traditional manufacturing, addressing the challenges of machining and manual methods in producing intricate structures. Additionally, 3D printing reduces production steps, enables mold-free manufacturing, and effectively saves time, reduces costs, and shortens production cycles.

In the aerospace industry, 3D printing has become a vital technology for producing complex components, offering significant advantages over traditional manufacturing methods. It enhances design freedom, enabling the creation of intricate geometries and internal structures that would be difficult or costly to produce using conventional techniques. Additionally, 3D printing reduces material waste, cuts manufacturing costs, and optimizes component designs. Selective Laser Melting (SLM), a widely used metal 3D printing technique, works by using a high-power laser to fully melt and fuse metallic powder particles layer by layer, enabling the production of fully dense, high-precision metal parts with complex geometries [133]. This method not only supports the fabrication of topology-optimized designs but also minimizes material waste, making it ideal for lightweight, high-performance aerospace components such as turbine blades or structural brackets. For example, in China's C919 aircraft, SLM technology developed by researchers at Northwestern Polytechnical University was used to produce 23 complex components, including the central winglet bar. Similarly, NASA employed SLA to manufacture a spacecraft fuel nozzle with a complex internal structure, reducing material waste and improving engine performance and reliability. Beyond component production, 3D printing offers greater flexibility in material processing, enabling the creation of metamaterials with extraordinary physical properties [134]. At NASA's Glenn Research Center, researchers used a model-based alloy design approach and laser-based rapid manufacturing technology to develop a new oxidation-dispersion-strengthened nickel-cobalt-based alloy, GRX-810. Compared to traditional alloys, GRX-810 demonstrated twice the strength, 1,000 times the creep resistance, and double the oxidation resistance at 1093 °C [135]. Similarly, researchers at the China Academy of Engineering Physics used direct-write 3D printing technology to create mechanical metamaterials with ultra-long mechanical adaptability and negative stiffness [136].

The application prospects of 3D printing in digital manufacturing have garnered significant attention. With continuous advancements in materials science, printing technology, and artificial intelligence (AI), 3D printing can now process not only traditional metals [137] and polymers [138] but also ceramics [139], composites [140], and even biological materials [141], greatly expanding its scope in intelligent manufacturing. AI plays a transformative role in optimizing the entire 3D printing process, from material composition [142] to structural design [136] and process optimization [143]. In material science, AI algorithms analyze vast datasets to predict and design novel materials with tailored properties, enhancing the performance of printed parts by identifying optimal material combinations and predicting their behavior under various conditions. In structural design, AI facilitates generative design, where algorithms propose and optimize complex geometries based on factors like strength, weight, and material usage, enabling the creation of highly efficient, performance-optimized structures [55]. Furthermore, AI drives process optimization by integrating intelligent control systems with 3D printing, enabling full automation from design to production [144]. AI continuously monitors real-time data during printing, making adjustments to printing parameters based on feedback, which allows for adaptive manufacturing that enhances precision, reduces defects, and minimizes errors [143]. This technological breakthrough, powered by AI, brings higher flexibility, sustainability, and precision to manufacturing, positioning 3D printing as a key driver in the transformation toward smart manufacturing.

4.5.2. AI-integrated digital twin

One of the main challenges in smart manufacturing is seamlessly integrating the physical and virtual spaces. Advances in data acquisition, communication, and computing technologies are making this integration increasingly feasible. Among these advancements, Digital Twin (DT) has emerged as a key enabler of cyber-physical integration, attracting significant interest from both academia and industry. DT integrates real-world and virtual data across the product lifecycle, developing advanced analytics to enhance decision-making and operational efficiency. Its applications span smart manufacturing, real-time system monitoring, and predictive maintenance. By continuously acquiring real-time data and updating virtual models, DT enables predictive adjustments, improving system reliability and efficiency. As a result, DT not only accelerates the adoption of smart manufacturing but also facilitates autonomous decision-making, real-time process optimization, and enhanced lifecycle management.

Digital Twin (DT) technology has significantly advanced materials science by enhancing the prediction of mechanical properties, optimizing additive manufacturing processes, and improving high-temperature material performance assessments. Researchers at the University of Science and Technology Beijing developed a DT-driven Hall-Petch model that integrates intrinsic material factors, such as valence electron distance and grain boundary energy, to eliminate empirical fitting constants and enable real-time optimization of polycrystalline metal properties [67]. In additive manufacturing, DT has been employed to refine process accuracy in laser powder bed fusion (L-PBF) of Inconel 718 alloy by integrating real-time monitoring and experimental data, thereby improving simulation reliability and optimizing the homogenization process. For instance, DT-assisted homogenization facilitated the dissolution of 67 % of the Laves phase within 0.5 h, ensuring more uniform microstructural evolution [145]. Furthermore, researchers at Northwestern Polytechnical University combined DT with Convolutional Neural Networks (CNN) to enhance the predictive modeling of creep fracture life in high-temperature titanium alloys, significantly improving model accuracy and adaptability through real-time feedback mechanisms [146]. These advancements underscore the transformative role of DT in materials research, enabling precise computational modeling, dynamic process optimization, and data-driven decision-making for the development of advanced materials.

Defect prediction and control are key to improving product quality in smart manufacturing. Researchers at Beihang University proposed a new DT-driven approach for End-face Defect Control (EF-DC) in hot-rolled coils [147]. Traditional EF-DC relies on physical-space data for defect identification, limiting its effectiveness. By integrating DT with deep learning, they created a multi-dimensional EF-DC model that maps surface quality across geometry, physics, behavior, and rules. This model combines operational and virtual data to track and correct defect-causing abnormalities in real time, enhancing defect detection and prevention. Additionally, they proposed a DT-based intelligent model for CNC machine tools that uses OPC interfaces to ensure interoperability between physical and digital spaces [148]. This model predicts potential failures, such as ball screw issues, by analyzing real-time data from vibration and temperature sensors, improving equipment reliability and efficiency.

With the continuous advancement of smart manufacturing technologies, DT has emerged as an important technology driving the transformation and upgrading of industrial production models. By enabling real-time data feedback and dynamic optimization of virtual models, T improves the precision and efficiency of manufacturing processes while providing new theoretical foundations and technological support for smart manufacturing.

4.5.3. AI agent-assisted material manufacturing

With the rapid development of LLM, emerging technologies represented by intelligent agents are set to significantly accelerate the shift from automation to smart manufacturing [149]. An agent is a computational system capable of autonomously perceiving its environment, making decisions, and executing tasks. Unlike traditional programs or algorithms, an agent exhibits autonomy, adaptability, collaboration, and learning abilities, allowing it to interact with its surroundings and adjust its behavior based on feedback, forming a closed-loop iterative system [150]. Agents typically gather environmental data via sensors, make decisions using reasoning mechanisms, and engage with the environment through actuators such as robots or control systems. The major characteristics of agents make them particularly useful in dynamic and complex environments, where they can effectively handle uncertainty and optimize decision-making processes.

Agents improve smart manufacturing at both local and global levels. At the local level, agents enhance quality control, process optimization, and equipment management [151,152]. Traditional quality checks rely on manual inspection or fixed machines, which can be slow and inaccurate. Agents use AI vision systems and sensors to detect defects automatically, improving accuracy. They also monitor equipment in real time, predicting failures to prevent breakdowns. Additionally, agents enable flexible production, adjusting machine settings to quickly switch between different products for customization. By improving quality, efficiency, and flexibility, agents make smart manufacturing faster, more reliable, and cost-effective. At the global level, agents can function as the "smart brain" of a factory, effectively integrating isolated departments, optimizing operational efficiency through real-time data exchange and intelligent scheduling, and reducing production costs. Through such global coordination, agents can achieve optimal allocation of production resources, enhancing the overall intelligence of the manufacturing process.

Although the empowerment of smart manufacturing through agents is still in its early stages, agents have already proven effective in advancing material chemistry research at the laboratory level [149]. Researchers from the Shanghai Institute of Ceramics, Chinese Academy of Sciences, inspired by brain functions, developed a multi-agent collaborative system with self-growth capabilities—Material Intelligence Scientist System (AI-Scientist). This system includes fundamental functional modules for completing complex material tasks, such as autonomous learning, data extraction, material reasoning, and interaction with the physical world. In scenarios with limited data, AI-Scientist demonstrated superior reasoning capabilities compared to Bayesian optimization through a dual-driven mode of data and knowledge, significantly improving the efficiency of material R&D. Notably, AI-Scientist independently optimized the performance of lithium-rich disordered cathode materials and, after four cycles, the materials produced in an automated laboratory based on the design reached a capacity of 282 mAh/g with 100 % capacity retention after 20 cycles, breaking through the Pareto frontier of the LiMnTiOF system and achieving world-leading results.

A preliminary case study of LLM enabled intelligent microstructure optimization and defects classification of welded titanium alloys is presented in Fig. 10, the technical roadmap of which consists of the LLM, auto-coding via AI agent, image processing, image mosaic, and ML for welding defect detection [114,153]. Based on the welding images obtained from high-throughput experiments, it is highlighted that the AI-integrated software is able to compile numerous images and extract crucial features from them, thereby efficiently facilitating the welding defect detection [114,153]. These achievements not only highlight the exceptional potential of agents in materials research but also lay the foundation for further developments in the field of smart manufacturing, signaling a new transformation in manufacturing through the widespread application of agent technology, with improvements in efficiency, process optimization, and the overcoming of innovation bottlenecks.

5. Prospect and challenge

The rapid advancement of AI is revolutionizing materials science, enabling unprecedented capabilities in data management, computational modeling, autonomous experimentation, and interdisciplinary collaboration. The establishment of a robust **materials data circulation infrastructure** is essential for harnessing the full potential of AI, ensuring seamless data integration, standardization, and accessibility across global research communities. **Advancements in AI power, algorithms, and tools** continue to drive the acceleration of material discovery and design, with deep learning, reinforcement learning, and generative models enhancing predictive accuracy and optimization efficiency. **Federated learning** presents a promising approach to training materials science large language models while preserving data privacy, fostering secure, large-scale collaboration. The development of **automated/intelligent experimental techniques** is streamlining high-throughput material synthesis and characterization, forming smart, closed-loop optimization processes that enhance R&D efficiency. Meanwhile, **multiscale modeling and digital twin technology** are breaking down spatial and temporal barriers, enabling real-time simulation, optimization, and precise control of materials manufacturing. Finally, fostering an **interdisciplinary workforce** with expertise in both AI and materials science will be critical for sustaining innovation and translating AI-driven advancements into practical industrial applications. Moving forward, the integration of AI with materials science will not only accelerate the discovery of next-generation materials but also drive sustainable, intelligent, and highly efficient manufacturing processes, shaping the future of materials innovation.

5.1. Materials data circulation infrastructure

For the future development of AI in materials science, it is essential to gather multi-scale, high-quality, and extensive materials data, including LLMs and the evolution of machine learning algorithms and software. The implementation of FAIR principles—Findability, Accessibility, Interoperability, and Reuse—across various materials communities, both in academia and industry, is becoming increasingly important. The construction of National materials big data center is poised to revolutionize how industries manage and utilize materials data in a globalized economy. Prospective benefits include improved data accessibility for researchers and manufacturers, which can lead to accelerated product development and more sustainable practices through better materials selection and usage. Future research should focus on developing an integrated big data cloud platform for automatic data collection, storage, mining, and application, building a diversified, multi-layered, and

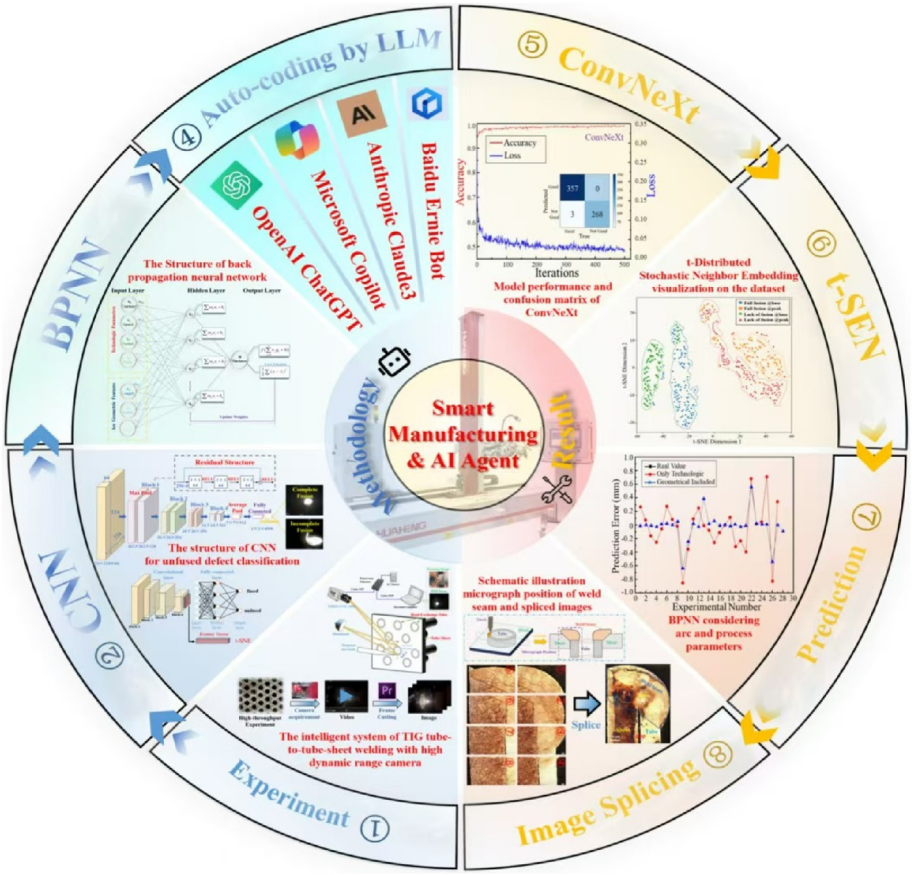

Fig. 10. Large language models enabled intelligent microstructure optimization and defects classification of welded titanium alloys [153].

systematic infrastructure for materials data. Establishing circulation technologies for material data identification, citation, evaluation, and trading will help overcome technical bottlenecks in data sharing, knowledge sharing, and information security. This will promote revolutionary changes in R&D models, accelerating the development and application of new materials.

Additionally, standardized protocols for data sharing can facilitate compliance with international regulations and enhance the interoperability of systems across borders, ensuring that data flows seamlessly between different jurisdictions and stakeholders. This includes incorporating common data dictionaries, data exchange formats, and pedigree standards. The absence of global or internationally recognized materials data standards for collection, storage, and exchange significantly hinders data sharing and community development in materials science. There is a pressing need for standardized materials data schemas, consistent file storage formats, international APIs for databases across different countries, and robust data exchange protocols between various software platforms and machines. As new methods and technologies are rapidly adopted, it becomes impossible to stay current with all the latest developments in the field. To address the diversity of datasets and machine learning methods, a checklist has been proposed to guide the assessment with substantial ML content [154]. However, it remains challenging to determine whether AI is ready for materials data standardization. Achieving informative, exhaustive, and adaptable data standardization in materials science continues to be an outstanding challenge.

5.2. Advances in AI power, algorithms, models, and tools

The continuous evolution of AI technologies is driving transformative changes in materials science, propelled by enhanced computational power, advanced algorithms, and the development of accessible toolkits. These advancements enable researchers to address previously insurmountable challenges, significantly accelerating material discovery and design.

State-of-the-art computational infrastructure underpins AI's integration into materials research. Specialized hardware—such as GPUs, high-performance computing (HPC) systems, and emerging quantum computing platforms—has dramatically increased the capacity to process large datasets and perform complex calculations. These computational advances are essential for supporting data-intensive AI models and enabling rapid simulation and analysis of materials.

Advancements in AI algorithms and models have expanded the capabilities of materials science beyond traditional methods. Machine learning approaches, initially dominated by regression and decision-tree models, have evolved to include deep learning, reinforcement learning, and generative models. These advanced models can predict material properties, optimize synthesis processes, and even design new materials with targeted characteristics. Notably, generative adversarial networks (GANs) and variational autoencoders (VAEs) are increasingly used to create materials with specific properties, such as enhanced strength or superconductivity.

The development of comprehensive AI toolkits and platforms has made these technologies more accessible to the materials science community. Platforms like the *Materials Project* and *AFLOW*, along with widely used machine learning libraries such as TensorFlow and PyTorch, enable seamless integration of AI into research workflows. These resources lower the barrier to entry, allowing researchers to leverage AI without extensive backgrounds in data science. Cloud-based services further facilitate the deployment of scalable models, offering on-demand computational resources that accelerate the design and optimization of new materials.

Looking ahead, four emerging AI-driven design patterns—reflection, tool use, planning, and multi-agent collaboration—are expected to drive significant progress. These approaches will enhance autonomous discovery, streamline complex workflows, and promote collaborative innovation, further accelerating advancements in materials science.

5.3. Federal training of materials scientific large language models

Federal training for materials scientific large language models marks a transformative shift in materials science. This decentralized approach enables privacy-preserving collaboration across disparate datasets, significantly accelerating material discovery, optimizing preparation processes, and enhancing material performance. Future trends in this field will likely emphasize integrating high-throughput simulations, AI-driven optimization, and multiscale modeling, which together will allow for the design of materials with highly tailored properties. The integration of data from experimental studies, real-time sensors, and large-scale simulations will not only expedite the discovery of novel materials but also enable more efficient preparation processes. Furthermore, federated models will facilitate the customization of materials for specific industrial needs, such as in aerospace, automotive, energy, and healthcare, fostering innovation and enhancing product functionality across sectors. The ability to optimize resource use and minimize waste aligns with the growing focus on sustainability, making this technology even more valuable in addressing global challenges.

The key advantage of federal training lies in its capacity to foster collaboration while protecting data privacy, allowing multiple stakeholders to contribute to model development without disclosing sensitive information. This decentralized framework supports the scalability of models and fosters global cooperation, facilitating the development of more robust and adaptable materials models. However, challenges remain, including the complexities of integrating diverse data sources, the computational expense of training large models, and ensuring model accuracy and adaptability to evolving materials data. Despite these hurdles, the continued evolution of federated learning in materials science holds immense potential to revolutionize industries, drive breakthroughs in material design, and promote sustainability across global manufacturing systems.

5.4. Intelligent experimental techniques

Future material experiments will rely on robotics technology, AI, and the integration of big data with cloud platform technologies to develop intelligent robotic systems for material experiments. These systems feature high sensitivity, multi-dimensional motion perception, multi-objective recognition, and intelligent decision-making feedback mechanisms. Deeply integrating intelligent robots with high-throughput material preparation, testing, and characterization technologies, they construct a physics platform with features of high-throughput, automation, and intelligence. Through multi-sensor real-time interaction technology and multi-scale modeling and simulation, the system is able to monitor changes in variables during experiments, providing timely feedback and optimizing the experimental plan. Through deep learning of experimental data and multi-objective optimization, the intelligent system can automatically adjust each experimental step and decision, forming an intelligent closed-loop optimization experimental plan aimed at achieving target performance in material research and development.

With a continuous upgrade of intelligent technologies, autonomous experimental systems may adopt a modular design. By developing modules tailored to different experimental scenarios, the system will match modules according to specific experimental needs, thus enabling full-process automation of various complex experimental tasks. This design supports complex experimental systems with multiple stages and parameters, making it adaptable to a wide range of application scenarios, including solid, liquid, and gas environments. With the enhancement of intelligence by using visual assistance technologies, the experimental platform can significantly improve the efficiency of intelligent robot operations and reduce ineffective experiments, further enhancing overall experimental efficiency. With the application of the metaverse and digital twins, experimental data will be presented in virtual reality (VR) and augmented reality (AR), allowing scientists to immerse themselves in and direct the experimental process. Moreover, the establishment of an intelligent cloud platform will become the core scheduling and control center of the entire intelligent experimental system, where all experimental data, control instructions, optimization algorithms, and intelligent decisions are centrally managed and intelligently scheduled. This networked and cloud-driven experimental model makes the storage, processing, and sharing of experimental data more efficient while also providing strong support for remote monitoring, collaborative optimization, and resource scheduling. This significantly enhances the system's scalability and flexibility, providing a convenient technological foundation for collaborative experiments across multiple fields and interdisciplinary cooperation. Moreover, when adopting blockchain technology, the data security and traceability will be greatly enhanced, ensuring the accuracy, transparency, and protection against tampering of data throughout the experimental process. Next-generation AI technologies will continuously drive the system's self-learning and optimization, enabling smarter, more automated experimental operations.

5.5. Multiscale modeling and digital twin technology

Future computational design of materials will focus on breaking spatial and temporal scale boundaries, developing cross-scale computational and autonomous methods and software for multiscale modeling. Machine learning molecular dynamics simulations have successfully achieved quantum mechanical precision simulations of millions of atoms, with the potential to truly transcend the simulation of spacetime boundaries, enabling the analysis of structure-property relationships at the micro-nano scale and the atomic-scale design of micro-nano device structures. Establishing on-demand, elastically scalable autonomous computational methods for materials will enable distributed, automated, and efficient computations. The integration of data-driven and computational materials engineering will break through bottlenecks in multiscale modeling and comprehensive computational design that encompass material composition, microstructure, macrostructure, properties, processing, and service behavior, achieving “pre-optimization” before material production or product manufacturing [114,130].

A substantial accumulation of foundational material data will be undertaken, covering multidimensional processing parameters, multi-field environments, multi-process data, and service behavior data. By integrating computational materials engineering and machine learning modeling, a digital model library and knowledge base will be constructed to dynamically describe the complex intrinsic relationships, interactions, and evolutionary behaviors of material composition, structure, processing, and performance. Breakthroughs in database technologies for the online collection, processing, storage, and computational interaction of multi-source heterogeneous data will enable real-time simulation visualization of material preparation and processing. Advances in physical and information system mapping and real-time interaction technologies will drive the construction of digital twin systems, enabling rapid online optimization and precise control of material manufacturing. These advancements will support the digitalization and intelligence of materials science and engineering.

5.6. Building an interdisciplinary workforce

As AI becomes increasingly integrated into materials science, there is a growing need for an interdisciplinary workforce that can bridge the gap between traditional materials science and cutting-edge AI technologies. This requires a combination of expertise in both domains, with professionals who can apply AI methods to solve complex materials problems. Materials scientists must gain proficiency in AI techniques, such as

machine learning, data analytics, and simulation tools, while AI practitioners need a strong understanding of the intricacies of materials science, including material properties, fabrication methods, and performance testing. Developing this cross-disciplinary skillset is essential to ensuring effective collaboration and innovation.

To address this need, educational programs must evolve, offering training that combines materials science with AI, data science, and computational methods. Collaborative initiatives between universities, research institutions, and industry partners can also help cultivate a workforce that is capable of addressing the challenges posed by AI integration in materials research. Efforts to promote such interdisciplinary training will ensure a future workforce capable of leveraging AI to its fullest potential, accelerating innovation and the development of new materials solutions. Building this skilled, collaborative workforce will be a critical factor in driving the next generation of materials science.

CRediT authorship contribution statement

Xue Jiang: Writing – original draft. **Dezhen Xue:** Writing – review & editing. **Yang Bai:** Writing – review & editing. **William Yi Wang:** Writing – review & editing. **Jianjun Liu:** Writing – review & editing. **Mingli Yang:** Writing – review & editing. **Yanjing Su:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is financially supported by the National Key Research and Development Program of China (2022YFB3707502), National Natural Science Foundation of China (92270001, 52201061), Guangdong Basic and Applied Basic Research Foundation (2023A1515140101).

References

[1] J. Xie, Y. Su, D. Xue, X. Jiang, H. Fu, Machine learning for materials research and development, Acta. Met. Sin. 57 (2021) 1343–1361.
[2] Materials Data Curation System, (n.d.). https://github.com/usnistgov/MDCS.
[3] S. Liu, Y. Su, H. Yin, D. Zhang, J. He, H. Huang, X. Jiang, X. Wang, H. Gong, Z. Li, H. Xiu, J. Wan, X. Zhang, An infrastructure with user-centered presentation data model for integrated management of materials data and services, npj Comput. Mater. 7 (2021) 88.
[4] H. Gong, J. He, X. Zhang, L. Duan, Z. Tian, W. Zhao, F. Gong, T. Liu, Z. Wang, H. Zhao, W. Jia, L. Zhang, X. Jiang, W. Chen, S. Liu, H. Xiu, W. Yang, J. Wan, A repository for the publication and sharing of heterogeneous materials data, Sci. Data 9 (2022) 787.
[5] G. Klimeck, M. McLennan, S.P. Brophy, G.B. Adams III, M.S. Lundstrom, nanohub.org: advancing education and research in nanotechnology, Comput. Sci. Eng. 10 (2008) 17–23.
[6] J.G. Kaufman, E.F. Begley, MatML: a data interchange markup language: this article updates and expands upon previously published information on MatML, a new extensible markup language designed specifically for the exchange of materials information. In particular, it describes the s, Adv. Mater. Process. 161 (2003) 35–37.
[7] C.W. Andersen, R. Armiento, E. Blokhin, G.J. Conduit, S. Dwaraknath, M.L. Evans, Á. Fekete, A. Gopakumar, S. Gražulis, A. Merkys, F. Mohamed, C. Oses, G. Pizzi, G. Rignanese, M. Scheidgen, L. Talirz, C. Toher, D. Winston, R. Aversa, K. Choudhary, P. Colinet, S. Curtarolo, D.D. Stefano, C. Draxl, S. Er, M. Esters, M. Fornari, M. Giantomassi, M. Govoni, G. Hautier, V. Hegde, M.K. Horton, P. Huck, G. Huhs, J. Hummelshøj, A. ariryaa, B. Kozinsky, S. Kumbhar, M. Liu, N. Marzari, A.J. Morris, A.A. Mostofi, K.A. Persson, G. Petretto, T. Purcell, F. Ricci, F. Rose, M. Scheffler, D. Speckhard, M. Uhrin, A. Vaitkus, P. Villars, D. Waroquiers, C. Wolverton, M. Wu, X. Yang, OPTIMADE, an API for exchanging materials data, Sci. Data 8 (2021) 217.
[8] Materials Genome Initiative for Global Competitiveness, (n.d.). https://www.mgi.gov/sites/default/files/documents/materials_genome_initiative-final.pdf.
[9] A. Agrawal, A. Choudhary, Perspective: materials informatics and big data: realization of the "fourth paradigm" of science in materials science, APL Mater. 4 (2016) 053208.
[10] D. Morgan, R. Jacobs, Opportunities and challenges for machine learning in materials science, Annu. Rev. Mater. Res. 50 (2020) 71–103.
[11] H.S. Rao, A. Mukherjee, Artificial neural networks for predicting the macromechanical behaviour of ceramic-matrix composites, Comput. Mater. Sci. 5 (1996) 307–322.
[12] X.-L. Yu, B. Yi, X.-Y. Wang, Prediction of the glass transition temperatures for polymers with artificial neural network, J. Theor. Comput. Chem. 7 (2008) 953–963.
[13] B. Meredig, A. Agrawal, S. Kirklin, J.E. Saal, J.W. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, Combinatorial screening for new materials in unconstrained composition space with machine learning, Phys. Rev. B 89 (2014) 94104.
[14] P. Raccuglia, K.C. Elbert, P.D.F. Adler, C. Falk, M.B. Wenny, A. Mollo, M. Zeller, S.A. Friedler, J. Schrier, A.J. Norquist, Machine-learning-assisted materials discovery using failed experiments, Nature 533 (2016) 73–76.
[15] J. Xiong, S.-Q. Shi, T.-Y. Zhang, A machine-learning approach to predicting and understanding the properties of amorphous metallic alloys, Mater. Des. 187 (2020) 108378.
[16] C. Wang, C. Shen, Q. Cui, C. Zhang, W. Xu, Tensile property prediction by feature engineering guided machine learning in reduced activation ferritic/martensitic steels, J. Nucl. Mater. 529 (2020) 151823.
[17] S. Lu, Q. Zhou, L. Ma, Y. Guo, J. Wang, Rapid discovery of ferroelectric photovoltaic perovskites and material descriptors via machine learning, Small Methods 3 (2019) 1900360.
[18] L. Ward, A. Dunn, A. Faghaninia, N.E.R. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K.A. Persson, G.J. Snyder, I. Foster, A. Jain, Matminer: an open source toolkit for materials data mining, Comput. Mater. Sci. 152 (2018) 60–69.
[19] J. Ma, B. Cao, S. Dong, Y. Tian, M. Wang, J. Xiong, S. Sun, MLMD: a programming-free AI platform to predict and design materials, npj Comput. Mater. 10 (2024) 59.
[20] S.P. Collins, T.D. Daff, S.S. Piotrkowski, T.K. Woo, Materials design by evolutionary optimization of functional groups in metal-organic frameworks, Sci. Adv. 2 (2016) e1600954.
[21] A. Seko, T. Maekawa, K. Tsuda, I. Tanaka, Machine learning with systematic density-functional theory calculations: application to melting temperatures of single-and binary-component solids, Phys. Rev. B 89 (2014) 54303.
[22] D. Xue, P.V. Balachandran, J. Hogden, J. Theiler, D. Xue, T. Lookman, Accelerated search for materials with targeted properties by adaptive design, Nat. Commun. 7 (2016) 1–9.
[23] D. Xue, P. V Balachandran, R. Yuan, T. Hu, X. Qian, E.R. Dougherty, T. Lookman, Accelerated search for $BaTiO_3$-based piezoelectrics with vertical morphotropic phase boundary using Bayesian learning, Proc. Natl. Acad. Sci. 113 (2016) 13301–13306.
[24] R. Yuan, Z. Liu, P. V Balachandran, D. Xue, Y. Zhou, X. Ding, J. Sun, D. Xue, T. Lookman, Accelerated discovery of large electrostrains in $BaTiO_3$-based piezoelectrics using active learning, Adv. Mater. 30 (2018) 1702884.
[25] Y. Tian, R. Yuan, D. Xue, Y. Zhou, X. Ding, J. Sun, T. Lookman, Role of uncertainty estimation in accelerating materials development via active learning, J. Appl. Phys. 128 (2020) 014103.
[26] Y. Tian, R. Yuan, D. Xue, Y. Zhou, Y. Wang, X. Ding, J. Sun, T. Lookman, Determining multi-component phase diagrams with desired characteristics using active learning, Adv. Sci. 8 (2021) 2003165.
[27] H. Zhang, H. Fu, S. Zhu, W. Yong, J. Xie, Machine learning assisted composition effective design for precipitation strengthened copper alloys, Acta Mater. 215 (2021) 117118.
[28] X. Yuan, M. Suvarna, J.Y. Lim, J. Pérez-Ramírez, X. Wang, Y.S. Ok, Active learning-based guided synthesis of engineered biochar for CO_2 capture, Environ. Sci. Technol. 58 (2024) 6628–6636.
[29] S. Xu, J. Li, P. Cai, X. Liu, B. Liu, X. Wang, Self-improving photosensitizer discovery system via Bayesian search with first-principle simulations, J. Am. Chem. Soc. 143 (2021) 19769–19777.
[30] J. Xiong, J.-C. He, X.-S. Leng, T.-Y. Zhang, Gaussian process regressions on hot deformation behaviors of FGH98 nickel-based powder superalloy, J. Mater. Sci. Technol. 146 (2023) 177–185.
[31] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.
[32] K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C.W. Park, A. Choudhary, A. Agrawal, S.J.L. Billinge, E. Holm, S.P. Ong, C. Wolverton, Recent advances and applications of deep learning methods in materials science, npj Comput. Mater. 8 (2022) 59.
[33] H. Li, Z. Wang, N. Zou, M. Ye, R. Xu, X. Gong, W. Duan, Y. Xu, Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation, Nat. Comput. Sci. 2 (2022) 367–377.
[34] X. Gong, S.G. Louie, W. Duan, Y. Xu, Generalizing deep learning electronic structure calculation to the plane-wave basis, Nat. Comput. Sci. 4 (2024) 752–760.
[35] Y. Wang, Y. Li, Z. Tang, H. Li, Z. Yuan, H. Tao, N. Zou, T. Bao, X. Liang, Z. Chen, S. Xu, C. Bian, Z. Xu, C. Wang, C. Si, W. Duan, Y. Xu, Universal materials model of deep-learning density functional theory Hamiltonian, Sci. Bull. 69 (2024) 2514–2521.
[36] Z. Tang, H. Li, P. Lin, X. Gong, G. Jin, L. He, H. Jiang, X. Ren, W. Duan, Y. Xu, A deep equivariant neural network approach for efficient hybrid density functional calculations, Nat. Commun. 15 (2024) 8815.
[37] M. Weng, Z. Wang, G. Qian, Y. Ye, Z. Chen, X. Chen, S. Zheng, F. Pan, Identify crystal structures by a new paradigm based on graph theory for building materials big data, Sci. China Chem. 62 (2019) 982–986.
[38] S. Li, Z. Chen, Z. Wang, M. Weng, J. Li, M. Zhang, J. Lu, K. Xu, F. Pan, Graph-based discovery and analysis of atomic-scale one-dimensional materials, Natl. Sci. Rev. 9 (2022) nwac028.

[39] W. Zhang, M. Weng, M. Zhang, Z. Chen, B. Wang, S. Li, F. Pan, Rapid mining of fast ion conductors via subgraph isomorphism matching, J. Am. Chem. Soc. 146 (2024) 18535–18543.

[40] L. Chen, B. Wang, W. Zhang, S. Zheng, Z. Chen, M. Zhang, C. Dong, F. Pan, S. Li, Crystal structure assignment for Unknown compounds from X-ray diffraction patterns with deep learning, J. Am. Chem. Soc. 146 (2024) 8098–8109.

[41] H. Zhang, Z. Wang, J. Ren, J. Liu, J. Li, Ultra-fast and accurate binding energy prediction of shuttle effect-suppressive sulfur hosts for lithium-sulfur batteries using machine learning, Energy Storage Mater. 35 (2021) 88–98.

[42] Z. Wang, H. Zhang, J. Ren, X. Lin, T. Han, J. Liu, J. Li, Predicting adsorption ability of adsorbents at arbitrary sites for pollutants using deep transfer learning, npj Comput. Mater. 7 (2021) 19.

[43] S. Wu, Z. Wang, H. Zhang, J. Cai, J. Li, Deep learning accelerates the discovery of two-dimensional catalysts for hydrogen evolution reaction, Energy & Environ. Mater. 6 (2023) e12259.

[44] D.P. Kingma, Auto-encoding Variational Bayes, 2013. ArXiv Prepr. ArXiv1312.6114.

[45] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (2020) 139–144.

[46] B.A. Sutton Rs, Reinforcement Learning: an Introduction, MIT Press, 1998.

[47] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, Adv. Neural Inf. Process. Syst. 33 (2020) 6840–6851.

[48] Z. Zhou, Y. Shang, X. Liu, Y. Yang, A generative deep learning framework for inverse design of compositionally complex bulk metallic glasses, npj Comput. Mater. 9 (2023) 15.

[49] Y. Shang, Z. Zhou, R. Han, H. Wang, X. Liu, Y. Yang, Customized design of amorphous solids by generative deep learning, Innov. Mater. 2 (2024) 100071.

[50] Z.Q. Chen, Y.H. Shang, X.D. Liu, Y. Yang, Accelerated discovery of eutectic compositionally complex alloys by generative machine learning, npj Comput. Mater. 10 (2024) 204.

[51] Y. Xian, P. Dang, Y. Tian, X. Jiang, Y. Zhou, X. Ding, J. Sun, T. Lookman, D. Xue, Compositional design of multicomponent alloys using reinforcement learning, Acta Mater. 274 (2024) 120017.

[52] D.X. Pengfei Dang, Jinlong Hu, Yuehui Xian, Cheng Li, Yumei Zhou, Xiangdong Ding, Jun Sun, Elastocaloric thermal battery: ultrahigh heat-storage capacity based on generative learning-designed phase-change alloys, Adv. Mater. (2025) 2412198.

[53] C. Zhong, J. Zhang, Y. Wang, Y. Long, P. Zhu, J. Liu, K. Hu, J. Chen, X. Lin, High-performance diffusion model for inverse design of high T_c superconductors with effective doping and accurate stoichiometry, InfoMat 6 (2024) e12519.

[54] C. Zhong, Y. Wang, Y. Long, J. Liu, K. Hu, J. Zhang, J. Chen, X. Lin, Enhancing superconductor critical temperature prediction: a novel machine learning approach integrating dopant recognition, ACS Appl. Mater. Interfaces 16 (2024) 60472–60481.

[55] B. Peng, Y. Wei, Y. Qin, J. Dai, Y. Li, A. Liu, Y. Tian, L. Han, Y. Zheng, P. Wen, Machine learning-enabled constrained multi-objective design of architected materials, Nat. Commun. 14 (2023) 6630.

[56] Y. Wang, W. Sha, M. Xiao, C.-W. Qiu, L. Gao, Deep-learning-enabled intelligent design of thermal metamaterials, Adv. Mater. 35 (2023) 2302387.

[57] C.M. Childs, N.R. Washburn, Embedding domain knowledge for machine learning of complex material systems, MRS Commun 9 (2019) 806–820.

[58] S. Kirklin, J.E. Saal, B. Meredig, A. Thompson, J.W. Doak, M. Aykol, S. Rühl, C. Wolverton, The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies, npj Comput. Mater. 1 (2015) 1–15.

[59] R. Yuan, Y. Tian, D. Xue, D. Xue, Y. Zhou, X. Ding, J. Sun, T. Lookman, Accelerated search for $BaTiO_3$-based ceramics with large energy storage at low fields using machine learning and experimental design, Adv. Sci. 6 (2019) 1901395.

[60] R. Yuan, D. Xue, D. Xue, J. Li, X. Ding, J. Sun, T. Lookman, Knowledge-based descriptor for the compositional dependence of the phase transition in BaTiO3-based ferroelectrics, ACS Appl. Mater. Interfaces 12 (2020) 44970–44980.

[61] Y. Zhang, C. Wen, C. Wang, S. Antonov, D. Xue, Y. Bai, Y. Su, Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models, Acta Mater. 185 (2020) 528–539.

[62] Y. Zhang, C. Wen, P. Dang, T. Lookman, D. Xue, Y. Su, Toward ultra-high strength high entropy alloys via feature engineering, J. Mater. Sci. Technol. 200 (2024) 243–252.

[63] R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, L.M. Ghiringhelli, SISSO: a compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates, Phys. Rev. Mater. 2 (2018) 83802.

[64] C.J. Bartel, C. Sutton, B.R. Goldsmith, R. Ouyang, C.B. Musgrave, L.M. Ghiringhelli, M. Scheffler, New tolerance factor to predict the stability of perovskite oxides and halides, Sci. Adv. 5 (2019) eaav0693.

[65] T. Wang, J. Hu, R. Ouyang, Y. Wang, Y. Huang, S. Hu, W.-X. Li, Nature of metal-support interaction for metal catalysts on oxide supports, Science 386 (2024) 915–920.

[66] C. Shen, C. Wang, X. Wei, Y. Li, S. van der Zwaag, W. Xu, Physical metallurgy-guided machine learning and artificial intelligent design of ultrahigh-strength stainless steel, Acta Mater. 179 (2019) 201–214.

[67] L. Jiang, H. Fu, H. Zhang, J. Xie, Physical mechanism interpretation of polycrystalline metals' yield strength via a data-driven method: a novel Hall–Petch relationship, Acta Mater. 231 (2022) 117868.

[68] S. Liu, T. Wen, A.S.L.S. Pattamatta, D.J. Srolovitz, A prompt-engineered large language model, deep learning workflow for materials classification, Mater. Today 80 (2024) 240–249.

[69] K. Hippalgaonkar, Q. Li, X. Wang, J.W. Fisher III, J. Kirkpatrick, T. Buonassisi, Knowledge-integrated machine learning for materials: lessons from gameplaying and robotics, Nat. Rev. Mater. 8 (2023) 241–260.

[70] M.C. Swain, J.M. Cole, ChemDataExtractor: a toolkit for automated extraction of chemical information from the scientific literature, J. Chem. Inf. Model. 56 (2016) 1894–1904.

[71] W. Wang, X. Jiang, S. Tian, P. Liu, D. Dang, Y. Su, T. Lookman, J. Xie, Automated pipeline for superalloy data by text mining, npj Comput. Mater. 8 (2022) 9.

[72] W. Wang, X. Jiang, S. Tian, P. Liu, T. Lookman, Y. Su, J. Xie, Alloy synthesis and processing by semi-supervised text mining, npj Comput. Mater. 9 (2023) 183.

[73] M.P. Polak, D. Morgan, Extracting accurate materials data from research papers with conversational language models and prompt engineering, Nat. Commun. 15 (2024) 1569.

[74] V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K.A. Persson, G. Ceder, A. Jain, Unsupervised word embeddings capture latent knowledge from materials science literature, Nature 571 (2019) 95–98.

[75] Z. Pei, J. Yin, P.K. Liaw, D. Raabe, Toward the design of ultrahigh-entropy alloys via mining six million texts, Nat. Commun. 14 (2023) 54.

[76] D.A. Boiko, R. MacKnight, B. Kline, G. Gomes, Autonomous chemical research with large language models, Nature 624 (2023) 570–578.

[77] Y. Kang, J. Kim, ChatMOF: an artificial intelligence system for predicting and generating metal-organic frameworks using large language models, Nat. Commun. 15 (2024) 4705.

[78] C. Kuenneth, R. Ramprasad, polyBERT: a chemical language model to enable fully machine-driven ultrafast polymer informatics, Nat. Commun. 14 (2023) 4099.

[79] S. Tian, X. Jiang, W. Wang, Z. Jing, C. Zhang, C. Zhang, T. Lookman, Y. Su, Steel design based on a large language model, Acta Mater. 285 (2025) 120663.

[80] J. He, X. Su, C. Wang, J. Li, Y. Hou, Z. Li, C. Liu, D. Xue, J. Cao, Y. Su, Machine learning assisted predictions of multi-component phase diagrams and fine boundary information, Acta Mater. 240 (2022) 118341.

[81] K. Terayama, R. Tamura, Y. Nose, H. Hiramatsu, H. Hosono, Y. Okuno, K. Tsuda, Efficient construction method for phase diagrams using uncertainty sampling, Phys. Rev. Mater. 3 (2019) 33802.

[82] Y. Zhang, D.W. Apley, W. Chen, Bayesian optimization for materials design with mixed quantitative and qualitative variables, Sci. Rep. 10 (2020) 4924.

[83] X. Pei, J. Pei, H. Hou, Y. Zhao, Optimizing casting process using a combination of small data machine learning and phase-field simulations, npj Comput. Mater. 11 (2025) 27.

[84] C. Kim, R. Batra, L. Chen, H. Tran, R. Ramprasad, Polymer design using genetic algorithm and machine learning, Comput. Mater. Sci. 186 (2021) 110067.

[85] P. Nikolaev, D. Hooper, N. Perea-Lopez, M. Terrones, B. Maruyama, Discovery of wall-selective carbon nanotube growth conditions via automated experimentation, ACS Nano 8 (2014) 10214–10222.

[86] P. Nikolaev, D. Hooper, F. Webber, R. Rao, K. Decker, M. Krein, J. Poleski, R. Barto, B. Maruyama, Autonomy in materials research: a case study in carbon nanotube growth, npj Comput. Mater. 2 (2016) 1–6.

[87] D.P. Tabor, L.M. Roch, S.K. Saikin, C. Kreisbeck, D. Sheberla, J.H. Montoya, S. Dwaraknath, M. Aykol, C. Ortiz, H. Tribukait, C. Amador-Bedolla, C.J. Brabec, B. Maruyama, K.A. Persson, A. Aspuru-Guzik, Accelerating the discovery of materials for clean energy in the era of smart automation, Nat. Rev. Mater. 3 (2018) 5–20.

[88] B.P. MacLeod, F.G.L. Parlane, T.D. Morrissey, F. Häse, L.M. Roch, K.E. Dettelbach, R. Moreira, L.P.E. Yunker, M.B. Rooney, J.R. Deeth, G.J. Ng, H. Situ, R.H. Zhang, M.S. Elliott, T.H. Haley, D.J. Dvorak, A. Aspuru-Guzik, J.E. Hein, C.P. Berlinguette, Self-driving laboratory for accelerated discovery of thin-film materials, Sci. Adv. 6 (2020) eaaz8867.

[89] B.P. MacLeod, F.G.L. Parlane, C.C. Rupnow, K.E. Dettelbach, M.S. Elliott, T.D. Morrissey, T.H. Haley, O. Proskurin, M.B. Rooney, N. Taherimakhsousi, D.J. Dvorak, H.N. Chiu, C.E.B. Waizenegger, K. Ocean, M. Mokhtari, C.P. Berlinguette, A self-driving laboratory advances the Pareto front for material properties, Nat. Commun. 13 (2022) 995.

[90] B. Burger, P.M. Maffettone, V. V Gusev, C.M. Aitchison, Y. Bai, X. Wang, X. Li, B.M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R.S. Sprick, A.I. Cooper, A mobile robotic chemist, Nature 583 (2020) 237–241.

[91] W. Zhang, M. Yu, T. Liu, M. Cong, X. Liu, H. Yang, Y. Bai, Q. Zhu, S. Zhang, H. Gu, X. Wu, Z. Zhang, Y. Wu, H. Tian, X. Li, W. Zhu, A.I. Copper, Accelerated discovery of molecular nanojunction photocatalysts for hydrogen evolution by using automated screening and flow synthesis, Nat. Synth. 3 (2024) 595–605.

[92] Q. Zhu, Y. Huang, D. Zhou, L. Zhao, L. Guo, R. Yang, Z. Sun, M. Luo, F. Zhang, H. Xiao, X. Tang, X. Zhang, T. Song, X. Li, B. Chong, J. Zhou, Y. Zhang, B. Zhang, J. Cao, G. Zhang, S. Wang, G. Ye, W. Zhang, H. Zhao, S. Cong, H. Li, L. Ling, Z. Zhang, W. Shang, J. Jiang, Y. Luo, Automated synthesis of oxygen-producing catalysts from Martian meteorites by a robotic AI chemist, Nat. Synth. 3 (2024) 319–328.

[93] R.W. Epps, M.S. Bowen, A.A. Volk, K. Abdel-Latif, S. Han, K.G. Reyes, A. Amassian, M. Abolhasani, Artificial chemist: an autonomous quantum dot synthesis bot, Adv. Mater. 32 (2020) 2001626.

[94] A. Slattery, Z. Wen, P. Tenblad, J. Sanjosé-Orduna, D. Pintossi, T. den Hartog, T. Noël, Automated self-optimization, intensification, and scale-up of photocatalysis in flow, Science 383 (2024) eadj1817.

[95] A.E. Gongora, B. Xu, W. Perry, C. Okoye, P. Riley, K.G. Reyes, E.F. Morgan, K.A. Brown, A Bayesian experimental autonomous researcher for mechanical design, Sci. Adv. 6 (2020) eaaz1708.

[96] P. Schwaller, B. Hoover, J.-L. Reymond, H. Strobelt, T. Laino, Extraction of organic chemistry grammar from unsupervised learning of chemical reactions, Sci. Adv. 7 (2021) eabe4166.

[97] N.J. Szymanski, B. Rendy, Y. Fei, R.E. Kumar, T. He, D. Milsted, M.J. McDermott, M. Gallant, E.D. Cubuk, A. Merchant, An autonomous laboratory for the accelerated synthesis of novel materials, Nature 624 (2023) 86–91.

[98] B. Mortazavi, X. Zhuang, T. Rabczuk, A. V Shapeev, Atomistic modeling of the mechanical properties: the rise of machine learning interatomic potentials, Mater. Horiz. 10 (2023) 1956–1968.

[99] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, Nature 559 (2018) 547–555.

[100] J. Behler, M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, Phys. Rev. Lett. 98 (2007) 146401.

[101] H. Wang, L. Zhang, J. Han, E. Weinan, DeePMD-kit: a deep learning package for many-body potential energy representation and molecular dynamics, Comput. Phys. Commun. 228 (2018) 178–184.

[102] W. Jia, H. Wang, M. Chen, D. Lu, L. Lin, R. Car, E. Weinan, L. Zhang, Pushing the limit of molecular dynamics with ab initio accuracy to 100 million atoms with machine learning, in: International Conference for High Performance Computing, Networking, Storage and Analysis, 2020, pp. 1–14, https://doi.org/10.1109/SC41405.2020.00009. Atlanta, GA, USA.

[103] Linfeng Zhang, (n.d.). https://awards.acm.org/award_winners/zhang_7608842.

[104] Z. Guo, D. Lu, Y. Yan, S. Hu, R. Liu, G. Tan, N. Sun, W. Jiang, L. Liu, Y. Chen, L. Zhang, M. Chen, H. Wang, W. Jia, Extending the limit of molecular dynamics with ab initio accuracy to 10 billion atoms, in: Proc. 27th ACM SIGPLAN Symp. Princ. Pract. Parallel Program, 2022, pp. 205–218.

[105] Z. Fan, Y. Wang, P. Ying, K. Song, J. Wang, Y. Wang, Z. Zeng, K. Xu, E. Lindgren, J.M. Rahm, A.J. Gabourie, J. Liu, H. Dong, J. Wu, Y. Chen, Z. Zhong, J. Sun, P. Erhart, Y. Su, T. Ala-Nissila, GPUMD: a package for constructing accurate machine-learned potentials and performing highly efficient atomistic simulations, J. Chem. Phys. 157 (2022) 114801.

[106] Y. Li, Z. Tang, Z. Chen, M. Sun, B. Zhao, H. Li, H. Tao, Z. Yuan, W. Duan, Y. Xu, Neural-network density functional theory based on variational energy minimization, Phys. Rev. Lett. 133 (2024) 76401.

[107] A.M. Kiirikki, H.S. Antila, L.S. Bort, P. Buslaev, F. Favela-Rosales, T.M. Ferreira, P.F.J. Fuchs, R. Garcia-Fandino, I. Gushchin, B. Kav, N. Kučerka, P. Kula, M. Kurki, A. Kuzmin, A. Lalitha, F. Lolicato, J.J. Madsen, M.S. Miettinen, C. Mingham, L. Monticelli, R. Nencini, A.M. Nesterenko, T.J. Piggot, Á. Piñeiro, N. Reuter, S. Samantray, F. Suárez-Lestón, R. Talandashti, O.H.S. Ollila, Overlay databank unlocks data-driven analyses of biomolecules for all, Nat. Commun. 15 (2024) 1136.

[108] N. Yao, X. Chen, Z.-H. Fu, Q. Zhang, Applying classical, ab initio, and machine-learning molecular dynamics simulations to the liquid electrolyte for rechargeable batteries, Chem. Rev. 122 (2022) 10970–11021.

[109] Q. Zhang, Y. Yuan, J. Zhang, P. Fang, J. Pan, H. Zhang, T. Zhou, Q. Yu, X. Zou, Z. Sun, F. Yan, Machine learning-aided design of highly conductive anion exchange membranes for fuel cells and water electrolyzers, Adv. Mater. 36 (2024) 2404981.

[110] D. Xu, Q. Zhang, X. Huo, Y. Wang, M. Yang, Advances in data-assisted high-throughput computations for material design, Mater. Genome Eng. Adv. 1 (2023) e11, https://doi.org/10.1002/mgea.11.

[111] A. Merchant, S. Batzner, S.S. Schoenholz, M. Aykol, G. Cheon, E.D. Cubuk, Scaling deep learning for materials discovery, Nature 624 (2023) 80–85.

[112] G. Wang, L. Peng, K. Li, L. Zhu, J. Zhou, N. Miao, Z. Sun, ALKEMIE: an intelligent computational platform for accelerating materials discovery and design, Comput. Mater. Sci. 186 (2021) 110064.

[113] X. Yang, Z. Wang, X. Zhao, J. Song, M. Zhang, H. Liu, MatCloud: a high-throughput computational infrastructure for integrated management of materials simulation, data and resources, Comput. Mater. Sci. 146 (2018) 319–333.

[114] W.Y. Wang, S. Zhang, G. Li, J. Lu, Y. Ren, X. Wang, X. Gao, Y. Su, H. Song, J. Li, Artificial intelligence enabled smart design and manufacturing of advanced materials: the endless Frontier in AI+ era, Mater. Genome Eng. Adv. 2 (2024) e56.

[115] W.Y. Wang, J. Li, W. Liu, Z.-K. Liu, Integrated computational materials engineering for advanced materials: a brief review, Comput. Mater. Sci. 158 (2019) 42–48.

[116] S.M. Hashemi, S. Parvizi, H. Baghbanijavid, A.T.L. Tan, M. Nematollahi, A. Ramazani, N.X. Fang, M. Elahinia, Computational modelling of process–structure–property–performance relationships in metal additive manufacturing: a review, Int. Mater. Rev. 67 (2022) 1–46.

[117] X. Wang, W. Xiong, Uncertainty quantification and composition optimization for alloy additive manufacturing through a CALPHAD-based ICME framework, npj Comput. Mater. 6 (2020) 188.

[118] J. Qi, X. Fan, D.I. Hoyos, M. Widom, P.K. Liaw, J. Poon, Integrated design of aluminum-enriched high-entropy refractory B2 alloys with synergy of high strength and ductility, Sci. Adv. 10 (2024) eadq0083.

[119] Y. Du, R. Schmid-Fetzer, J. Wang, S. Liu, J. Wang, Z. Jin, Computational Design of Engineering Materials: Fundamentals and Case Studies, Cambridge University Press, 2023.

[120] W.Y. Wang, B. Tang, D. Lin, C. Zou, Y. Zhang, S.-L. Shang, Q. Guan, J. Gao, L. Fan, H. Kou, H. Song, J. Ma, X. Hui, M.C. Gao, Z. Liu, J. Li, A brief review of data-driven ICME for intelligently discovering advanced structural metal materials: insight into atomic and electronic building blocks, J. Mater. Res. 35 (2020) 872–889.

[121] V. Attari, D. Khatamsaz, D. Allaire, R. Arroyave, Towards inverse microstructure-centered materials design using generative phase-field modeling and deep variational autoencoders, Acta Mater. 259 (2023) 119204.

[122] W.Y. Wang, P. Li, D. Lin, B. Tang, J. Wang, Q. Guan, Q. Ye, H. Dai, J. Gao, X. Fan, H. Kou, H. Song, F. Zhou, J. Ma, Z. Liu, J. Li, W. Liu, DID code: a bridge connecting the materials genome engineering database with inheritable integrated intelligent manufacturing, Engineering 6 (2020) 612–620.

[123] M. Umehara, H.S. Stein, D. Guevarra, P.F. Newhouse, D.A. Boyd, J.M. Gregoire, Analyzing machine learning models to accelerate generation of fundamental materials insights, npj Comput. Mater. 5 (2019) 34.

[124] P.L. Craig, Modeling Software for Materials 4.0, 2021.

[125] W.-Y. Wang, J. Yin, Z. Chai, X. Chen, W. Zhao, J. Lu, F. Sun, Q. Jia, X. Gao, B. Tang, X. Hui, H. Song, F. Xue, Z. Liu, J. Li, Big data-assisted digital twins for the smart design and manufacturing of advanced materials: from atoms to products, J. Mater. Inf. 2 (2022) 1–27, https://doi.org/10.20517/jmi.2021.11.

[126] Y. Yi, Y. Yan, X. Liu, Z. Ni, J. Feng, J. Liu, Digital twin-based smart assembly process design and application framework for complex products and its case study, J. Manuf. Syst. 58 (2021) 94–107.

[127] Y. Xu, X. Liu, X. Cao, C. Huang, E. Liu, S. Qian, X. Liu, Y. Wu, F. Dong, C.-W. Qiu, J. Qiu, K. Hua, W. Su, J. Wu, H. Xu, Y. Han, C. Fu, Z. Yin, M. Liu, R. Roepman, S. Dietmann, M. Virta, F. Kengara, Z. Zhang, L. Zhang, T. Zhao, J. Dai, J. Yang, L. Lan, M. Luo, Z. Liu, T. An, B. Zhang, X. He, S. Cong, X. Liu, W. Zhang, J.P. Lewis, J.M. Tiedje, Q. Wang, Z. An, F. Wang, L. Zhang, T. Huang, C. Lu, Z. Cai, F. Wang, J. Zhang, Artificial intelligence: a powerful paradigm for scientific research, Innovation 2 (2021) 100179, https://doi.org/10.1016/j.xinn.2021.100179.

[128] N. Nicola, Can artificial intelligence create the next wonder material, Nature 533 (2016) 22–25.

[129] A.R. Oganov, C.J. Pickard, Q. Zhu, R.J. Needs, Structure prediction drives materials discovery, Nat. Rev. Mater. 4 (2019) 331–348.

[130] X. Zhang, L. Wang, J. Helwig, Y. Luo, C. Fu, Y. Xie, M. Liu, Y. Lin, Z. Xu, K. Yan, K. Adams, M. Weiler, X. Li, T. Fu, Y. Wang, A. Strasser, H. Yu, Y. Xie, X. Fu, S. Xu, Y. Liu, Y. Du, A. Saxton, H. Ling, H. Lawrence, H. Stärk, S. Gui, C. Edwards, N. Gao, A. Ladera, T. Wu, E.F. Hofgard, A.M. Tehrani, R. Wang, A. Daigavane, M. Bohde, J. Kurtin, Q. Huang, T. Phung, M. Xu, C.K. Joshi, S.V. Mathis, K. Azizzadenesheli, A. Fang, A. Aspuru-Guzik, e. Bekkers, M. Bronstein, M. Zitnik, A. Anandkurmar, S. Ermon, P. Liò, R. Yu, S. Günnemann, J. Leskovec, H. Ji, J. Sun, R. Barzilay, T. Jaakkola, C.W. Coley, X. Qian, X. Qian, T. Smidt, S. Ji, Artificial Intelligence for Science in Quantum, Atomistic, and Continuum Systems, 2023. ArXiv Prepr. ArXiv2307.08423.

[131] H. Wang, T. Fu, Y. Du, W. Gao, K. Huang, Z. Liu, P. Chandak, S. Liu, P. Van Katwyk, A. Deac, A. Anandkumar, K. Bergen, C.P. Gomes, S. Ho, P. Kohli, J. Lasenby, J. Leskovec, T. Liu, A. Manrai, D. Marks, B. Ramsundar, L. Song, J. Sun, J. Tang, P. Veličković, M. Welling, L. Zhang, C.W. Coley, Y. Vengio, M. Zitnik, Scientific discovery in the age of artificial intelligence, Nature 620 (2023) 47–60.

[132] National Science and Technology Council, Materials Genome Initiative Strategic Plan, 2021.

[133] X. Gao, X. Lin, Q. Yan, Z. Wang, X. Yu, Y. Zhou, Y. Hu, W. Huang, Effect of Cu content on microstructure and mechanical properties of in-situ β phases reinforced Ti/Zr-based bulk metallic glass matrix composite by selective laser melting (SLM), J. Mater. Sci. Technol. 67 (2021) 174–185.

[134] M.Y. Khalid, Z.U. Arif, A. Tariq, M. Hossain, R. Umer, M. Bodaghi, 3D printing of active mechanical metamaterials: a critical review, Mater. Des. (2024) 113305.

[135] T.M. Smith, C.A. Kantzos, N.A. Zarkevich, B.J. Harder, M. Heczko, P.R. Gradl, A.C. Thompson, M.J. Mills, T.P. Gabb, J.W. Lawson, A 3D printable alloy designed for extreme environments, Nature 617 (2023) 513–518.

[136] S. Yu, H. Chai, Y. Xiong, M. Kang, C. Geng, Y. Liu, Y. Chen, Y. Zhang, Q. Zhang, C. Li, H. Wei, Y. Zhao, F. Yu, A. Lu, Studying complex evolution of hyperelastic materials under external field stimuli using artificial neural networks with spatiotemporal features in a small-scale dataset, Adv. Mater. 34 (2022) 2200908.

[137] Z. Liu, M. Ding, J. Xie, Advancements in digital manufacturing for metal 3D printing, Acta. Met. Sin. 60 (2023) 569–584.

[138] E. Liao, Y. Zhang, Y. Su, C. Zhang, C. Geng, C. Li, X. Liu, Y. Liu, A. Lu, Active mixing of two-component viscoelastic silicone ink at molecular level for spatiotemporally controlled 3D/4D printing of cellular silicones, Polymer (Guildf). 302 (2024) 127053.

[139] F. Zhang, Z. Li, M. Xu, S. Wang, N. Li, J. Yang, A review of 3D printed porous ceramics, J. Eur. Ceram. Soc. 42 (2022) 3351–3373.

[140] T. Ma, Y. Zhang, K. Ruan, H. Guo, M. He, X. Shi, Y. Guo, J. Kong, J. Gu, Advances in 3D printing for polymer composites: a review, InfoMat 6 (2024) e12568.

[141] Z. Ma, J. Wang, L. Qin, A. Chortos, 3D printing in biofabrication: from surface textures to biological engineering, Chem. Eng. J. (2024) 156477.

[142] Z. Zhu, D.W.H. Ng, H.S. Park, M.C. McAlpine, 3D-printed multifunctional materials enabled by artificial-intelligence-assisted fabrication technologies, Nat. Rev. Mater. 6 (2021) 27–47.

[143] Z. Ren, L. Gao, S.J. Clark, K. Fezzaa, P. Shevchenko, A. Choi, W. Everhart, A.D. Rollett, L. Chen, T. Sun, Machine learning–aided real-time detection of keyhole pore generation in laser powder bed fusion, Science 379 (2023) 89–94.

[144] J.A. Lee, J. Park, M.J. Sagong, S.Y. Ahn, J.-W. Cho, S. Lee, H.S. Kim, Active learning framework to optimize process parameters for additive-manufactured Ti-6Al-4V with high strength and ductility, Nat. Commun. 16 (2025) 931.

[145] M. Chen, Q. Du, R. Shi, H. Fu, Z. Liu, J. Xie, Phase field simulation of microstructure evolution and process optimization during homogenization of additively manufactured Inconel 718 alloy, Front. Mater. 9 (2022) 1043249.

[146] B. Zong, J. Li, C. Zhou, P. Wang, B. Tang, R. Yuan, Enhancing creep rupture life prediction of high-temperature titanium alloys using convolutional neural networks, Mater, Genome Eng. Adv. 2 (2024) e68.

[147] F. Xiang, S. Zhou, Y. Zuo, F. Tao, Digital twin driven end-face defect control method for hot-rolled coil with cloud-edge collaboration, IEEE Trans. Ind. Inf. 19 (2022) 1674–1682, https://doi.org/10.1109/TII.2022.3188508.

[148] W. Luo, T. Hu, W. Zhu, F. Tao, Digital twin modeling method for CNC machine tool, in: 2018 IEEE 15th Int. Conf. Networking, Sens. Control, 2018, pp. 1–4.

[149] S. Yu, N. Ran, J. Liu, Large-language models: the game-changers for materials science research, Artif. Intell. Chem. 2 (2024) 100076.

[150] T. Song, M. Luo, X. Zhang, L. Chen, Y. Huang, J. Cao, Q. Zhu, D. Liu, B. Zhang, G. Zou, G. Zhang, F. Zhang, W. Shang, Y. Fu, J. Jiang, Y. Luo, A multiagent-driven robotic AI chemist enabling autonomous chemical research on demand, J. Am. Chem. Soc. (2024), https://doi.org/10.1021/jacs.4c17738.

[151] M.H. Prince, H. Chan, A. Vriza, T. Zhou, V.K. Sastry, Y. Luo, M.T. Dearing, R.J. Harder, R.K. Vasudevan, M.J. Cherukara, Opportunities for retrieval and tool augmented large language models in scientific facilities, npj Comput. Mater. 10 (2024) 251.

[152] Y. Ruan, C. Lu, N. Xu, Y. He, Y. Chen, J. Zhang, J. Xuan, J. Pan, Q. Fang, H. Gao, X. Shen, N. Ye, Q. Zhang, Y. Mo, An automatic end-to-end chemical synthesis development platform powered by large language models, Nat. Commun. 15 (2024) 10160.

[153] S. Zhang, W.Y. Wang, X. Wang, G. Li, Y. Ren, X. Gao, F. Sun, B. Tang, H. Song, J. Li, Large language models enabled intelligent microstructure optimization and defects classification of welded titanium alloys, J. Mater. Inforatics 4 (2024) 34, https://doi.org/10.20517/jmi.2024.64.

[154] K.T. Butler, K. Choudhary, G. Csanyi, A.M. Ganose, S. V Kalinin, D. Morgan, Setting standards for data driven materials science, npj Comput. Mater. 10 (2024) 231.