
A Survey on the Integration of Robotics AI and Autonomous Laboratories in Materials Science and Chemistry

www.surveyx.cn

Abstract

This survey paper explores the transformative integration of robotics, artificial intelligence (AI), and autonomous laboratories in materials science and chemistry. The paper highlights the role of advanced technologies such as large language models (LLMs) and machine learning (ML) in enhancing research methodologies, improving efficiency, and accelerating materials discovery. Self-Driving Laboratories (SDLs) exemplify this transformation by optimizing experimental workflows through closed-loop systems. The integration of high-throughput screening techniques with ML-assisted diagnostics has achieved substantial throughput improvements, underscoring the potential of AI-driven methodologies in advancing materials science. Despite these advancements, challenges such as data quality, model interpretability, and scalability persist, necessitating ongoing research and interdisciplinary collaboration. Future directions focus on improving computational methods, expanding ML frameworks, and integrating robust experimental data to fully harness the potential of AI and robotics in materials discovery. By addressing these challenges, the field can achieve significant progress in the discovery and development of novel materials, paving the way for future innovations and breakthroughs.

1 Introduction

1.1 Context and Motivation

The integration of robotics and artificial intelligence (AI) in materials science and chemistry is essential for navigating the extensive chemical space and addressing the limitations of traditional methods [1]. Conventional approaches often struggle with interpreting complex visual and textual data, especially concerning bio-inspired materials, which necessitates advanced analytical techniques [2]. This integration aims to enhance the prediction of material properties through data-driven methods, providing a more efficient alternative to traditional experimental techniques [3].

Deep learning techniques significantly contribute to the discovery and characterization of new crystalline materials by revealing intricate structure-property relationships that conventional methods may overlook [4]. The search for new transparent conducting materials (TCMs) is particularly challenged by existing data-driven methods, often limited by inadequate datasets [5]. Furthermore, there is a pressing need for methodologies that can effectively manage inhomogeneous measurement noise and anisotropic kernels, thus enhancing the efficiency and accuracy of autonomous materials discovery [6].

Current applications of machine learning (ML) in chemistry and materials science highlight significant research gaps that must be addressed to maximize ML's impact in these fields [7]. The integration of robotics and AI is critical for bridging these gaps, enabling the efficient discovery and development of new materials and fostering innovative research methodologies. Collectively, these efforts aim to

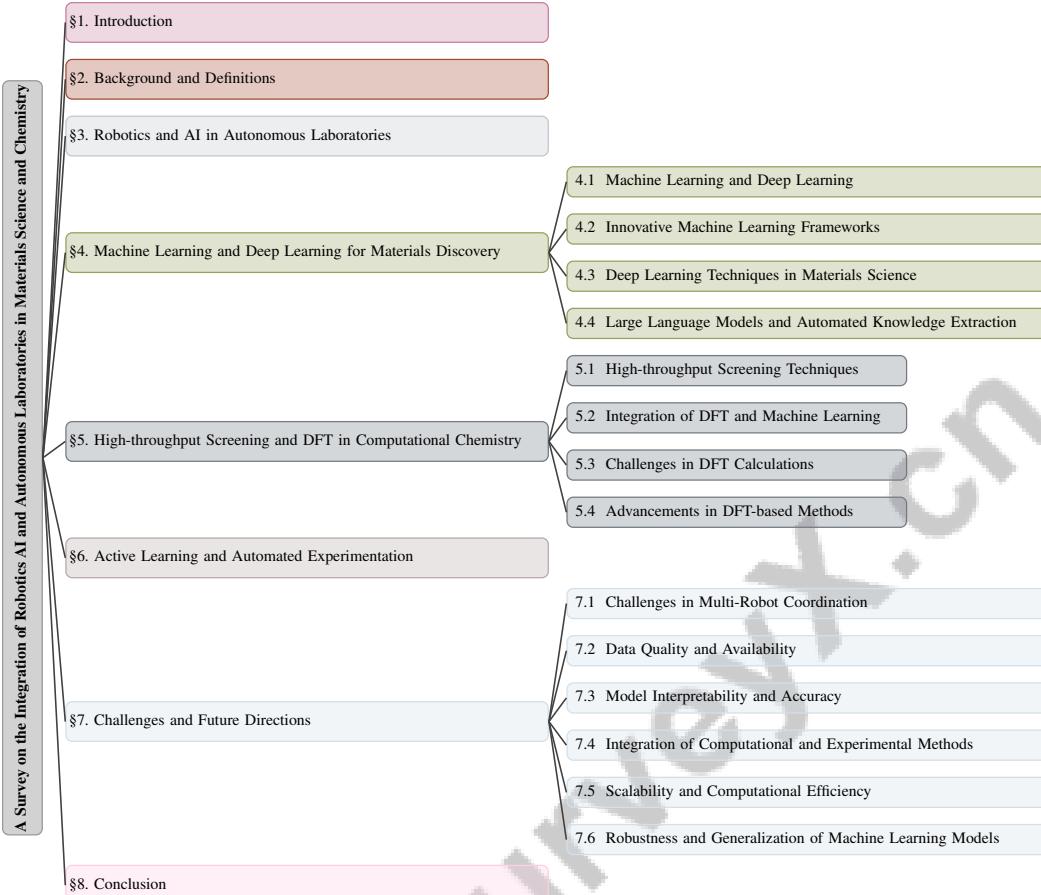


Figure 1: chapter structure

transform research outcomes, equipping the materials science community to tackle the complexities of chemical and materials research.

1.2 Transformative Impact of AI and Automation

The incorporation of artificial intelligence (AI) and automation into materials science and chemistry is set to revolutionize research methodologies by addressing core challenges in discovering and developing high-quality materials. The LLMatDesign framework exemplifies this potential, utilizing large language models (LLMs) to interpret human instructions and facilitate materials discovery, thereby enhancing research efficiency [1]. Data-driven methods and deep learning techniques have made significant strides in crystalline materials discovery, underscoring the profound impact of AI and automation on scientific advancement [4].

Innovative AI workflows that integrate LLMs, Bayesian optimization, and active learning loops demonstrate the potential to enhance catalyst synthesis efficiency, streamline synthetic routes, and improve the prediction accuracy of material properties [8]. The Intelligent Extraction Large Model (IELM) further automates knowledge extraction, minimizing the need for manual annotation and significantly improving upon traditional methods [9].

In structure-property discovery, a Bayesian deep learning framework autonomously directs exploration in scanning tunneling microscopy (STM) measurements, optimizing material properties and enhancing discovery efficiency [10]. This approach highlights the importance of AI-driven solutions in improving the coherence and productivity of research workflows within the chemical sciences [11].

Despite these advancements, challenges persist in fully leveraging machine learning (ML) capabilities in chemistry and materials science. The underutilization of ML for advancing scientific understanding and practical applications is a critical issue requiring ongoing efforts for effective integration [7].

Establishing a robust data ecosystem that supports machine learning is vital for facilitating model comparisons and advancing the field through improved access to data and methodologies [12].

Recent advancements in AI and automation are fundamentally transforming research methodologies and outcomes in materials science and chemistry. These innovations not only accelerate scientific inquiry but also enhance data extraction and analysis from vast unstructured literature. AI-driven tools can now identify correlations between material properties and synthesis processes without relying on pre-existing knowledge models, as illustrated by AlphaFold's deep learning applications for predicting protein structures. Additionally, natural language processing (NLP) frameworks are being employed to systematically extract and categorize material information from extensive scientific literature, enabling researchers to track trends and facilitate knowledge discovery. Collectively, these innovations herald a new paradigm in materials science and chemistry, effectively harnessing the synergy between human creativity and machine intelligence to foster more efficient and informed research practices [13, 14, 15, 16].

1.3 Interdisciplinary Approach

The integration of robotics, artificial intelligence (AI), and autonomous laboratories within materials science and chemistry exemplifies an interdisciplinary approach crucial for fostering innovation and accelerating scientific discovery. This convergence necessitates collaboration across diverse fields, including materials science, computer science, and data science, to effectively harness the potential of these advanced technologies [17]. By bridging these disciplines, researchers can leverage the strengths of each to address complex challenges in materials discovery and development.

In this interdisciplinary context, robotics and automation provide the framework for conducting experiments and collecting data with unprecedented precision and efficiency. AI and machine learning algorithms have emerged as advanced tools for data analysis, significantly enhancing the extraction of valuable insights from extensive datasets across various scientific fields. These technologies facilitate the identification of correlations between research features without relying on pre-existing knowledge, enabling new discoveries in materials science and other domains. For instance, AI applications like AlphaFold utilize deep learning to predict protein structures by analyzing large datasets, illustrating the transformative potential of AI in automating data processing and uncovering novel scientific findings [16, 15, 18, 19]. This synergy enables exploration of vast chemical spaces and prediction of material properties, traditionally time-consuming and resource-intensive tasks.

Furthermore, the integration of large language models (LLMs) and other AI-driven methodologies enhances the interpretation and processing of textual and visual data, broadening the scope of materials research. These technologies enable innovative workflows that streamline experimental processes, reduce human intervention, and improve reproducibility. The interdisciplinary approach accelerates discovery and enhances the quality and reliability of research outcomes by leveraging advanced techniques such as natural language processing (NLP) and deep learning to automate knowledge extraction from vast amounts of unstructured scientific literature. This allows researchers to efficiently identify and analyze critical information, trends, and properties of materials, ultimately facilitating informed experimentation and innovation across various domains, including materials science and chemistry [16, 11, 20, 13, 18].

Collaborative efforts among different scientific domains are essential for overcoming the limitations of traditional methods and driving advancements in materials science and chemistry. Cultivating a culture of interdisciplinary collaboration facilitates groundbreaking innovations and addresses critical challenges in the scientific community, such as the need for improved explainability, ethical considerations, and environmental efficiency. This collaborative approach is vital for navigating contemporary research complexities, as evidenced by recent advancements in AI that have transformed computing education and methodologies for knowledge extraction in materials science. Such innovations have the potential to significantly enhance the development and application of autonomous systems, ultimately revolutionizing various scientific fields [21, 22, 23, 18].

1.4 Structure of the Survey

The survey is meticulously organized to provide a comprehensive exploration of the integration of robotics, artificial intelligence (AI), and autonomous laboratories within materials science and chemistry. It begins with an **Introduction**, which highlights the significance of this interdisciplinary

approach in accelerating materials discovery and development. The **Context and Motivation** section emphasizes the necessity for advanced methodologies to navigate the complexities of chemical spaces. The **Transformative Impact of AI and Automation** elucidates the revolutionary potential of these technologies in enhancing research methodologies. Furthermore, the **Interdisciplinary Approach** underscores the necessity of cross-disciplinary collaboration to harness the full potential of advanced technologies [17].

Following the introduction, the **Background and Definitions** section provides an overview of the core concepts and technologies involved, such as robotics, AI, large language models (LLMs), machine learning, deep learning, high-throughput screening, and density functional theory (DFT). This foundational knowledge sets the stage for a deeper understanding of the subsequent sections.

The survey then progresses to **Robotics and AI in Autonomous Laboratories**, discussing the utilization of robotics and AI in autonomous laboratory settings. This section explores the benefits of automation in experimentation, data collection, and analysis. The **Role of Self-Driving Labs (SDLs)** is examined, highlighting their impact on research efficiency [9]. The **Integration of Robotics and AI for Experimental and Computational Synergy** and **Automation in Sample Preparation and Analysis** are also explored, illustrating improvements in efficiency and precision.

In **Machine Learning and Deep Learning for Materials Discovery**, the application of these techniques in predicting material properties and exploring chemical spaces is examined. This section delves into innovative machine learning frameworks and their contributions to materials discovery, as well as the role of deep learning techniques in advancing research.

The **High-throughput Screening and DFT in Computational Chemistry** section explores the integration of high-throughput screening techniques and DFT, focusing on the rapid assessment and identification of promising candidates. It discusses the synergy between DFT and machine learning, challenges in DFT calculations, and recent advancements in DFT-based methods.

Next, **Active Learning and Automated Experimentation** highlights the role of active learning in optimizing experimental workflows and the importance of automated experimentation in achieving reproducible research outcomes. This section also covers real-time monitoring and anomaly detection to improve experimental accuracy.

The survey concludes with a section titled **Challenges and Future Directions**, which identifies significant obstacles in the integration of advanced technologies, such as natural language processing (NLP) and large language models (LLMs), within materials science and chemistry. It highlights issues like the limited effectiveness of current LLMs in understanding complex materials science concepts, the scarcity of high-quality datasets for training models, and the need for standardized guidelines to facilitate the extraction of structured data from unstructured scientific literature. The section outlines a roadmap for overcoming these challenges, emphasizing the importance of developing domain-specific models and multi-modal datasets to enhance materials discovery and innovation [16, 24, 20, 13, 25]. It discusses potential future directions and innovations, addressing issues such as multi-robot coordination, data quality, model interpretability, and scalability.

Finally, the **Conclusion** summarizes the key points discussed in the paper, emphasizing the transformative potential of integrating robotics, AI, and autonomous laboratories in advancing materials science and chemistry. The significance of ongoing research and collaboration in this interdisciplinary field is highlighted, as it not only facilitates the extraction of critical knowledge from extensive literature through advanced natural language processing techniques but also fosters innovative solutions and breakthroughs in material synthesis, characterization, and application. This collaborative effort is essential for harnessing the vast amounts of unstructured data available, enabling researchers to efficiently identify trends, optimize experimental conditions, and ultimately accelerate the discovery of novel materials [15, 13, 18, 16]. The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Robotics and Automation in Materials Science

Robotics and automation are pivotal in advancing materials science by enhancing experimental efficiency, precision, and reproducibility. High-throughput techniques, especially in synthesizing perovskite-inspired materials, significantly improve synthesis and characterization processes, as

demonstrated by accelerated photovoltaic material development [26]. The evolution of materials acceleration platforms (MAPs), such as AMANDA, illustrates robotics' capability to automate complex tasks and expedite materials discovery [27]. Automated systems like thin-layer chromatography (TLC) analysis streamline processes, standardize experiments, and facilitate rapid material discovery [28]. High-throughput first principles calculations, essential for mapping compositional phase diagrams, address traditional methods' inefficiencies and enhance exploration of vast compositional spaces [29]. The AFLOW framework sets a benchmark for automating these calculations, offering tools for data verification and dissemination [30].

The synergy between robotics and AI further accelerates materials science innovation. The SARA method combines robotic synthesis with AI to efficiently explore synthesis phase spaces and characterize materials [31]. The AI-driven inverse design system (AIDS) automates material design by predicting properties and generating new chemical structures, streamlining novel materials discovery [32]. Machine learning interatomic potentials (MLPs) revolutionize computational materials science by enabling simulations of large systems with ab initio accuracy [33].

Challenges persist, including scarcity of annotated materials data and complexity of hierarchical materials structures, necessitating robust representations to capture materials science nuances [17]. The exponential growth of potential materials due to system complexity renders traditional methods impractical [34]. The lack of embodied intelligence in current robotic systems limits interaction with dynamic environments, posing significant challenges for future developments [35]. Addressing these challenges is crucial for advancing robotics and automation, paving the way for intelligent robotic systems capable of active participation in control processes akin to biological systems [36]. Overcoming these hurdles will enable continued integration of robotics and automation, driving transformative innovations in the field.

2.2 Artificial Intelligence and Large Language Models (LLMs)

AI and LLMs are increasingly pivotal in transforming materials science and chemistry by providing advanced tools for data analysis, hypothesis generation, and automated experimentation. AI integration enhances predictive accuracy and efficiency in materials discovery by processing vast datasets and uncovering complex patterns [37]. The LLMatDesign framework exemplifies this integration, autonomously discovering materials by interpreting user-defined objectives without extensive datasets, highlighting AI's potential to streamline discovery workflows [1]. Domain-specific LLMs, such as Polymetis, utilize a dataset of approximately 2 million material knowledge instructions, demonstrating these models' ability to provide accurate, structured responses tailored to materials science [9]. The Cephalo series of multimodal vision large language models (V-LLMs) integrates visual and linguistic data, enhancing understanding and interpretation of complex materials science problems [2].

Despite advancements, significant challenges persist. The scarcity of high-quality, annotated data necessary for training machine learning models limits AI's predictive accuracy and applicability in materials science [8]. This challenge is compounded by the need for robust benchmarks, such as MaCBench, which assess vision-language models' capabilities in real-world chemistry and materials science tasks, focusing on data extraction, experimental understanding, and results interpretation [38]. Integrating LLMs into compound AI systems for end-to-end optimization requires sophisticated training-time optimization methods to fully harness their potential [39]. Trust in AI automation agents, particularly LLMs, is crucial, emphasizing the importance of reliability, openness, and task-specific characteristics for broader acceptance and integration [40].

3 Robotics and AI in Autonomous Laboratories

The integration of robotics and AI has revolutionized autonomous laboratories, exemplified by Self-Driving Laboratories (SDLs). These laboratories minimize human intervention, enhance operational efficiency, and utilize advanced technologies to optimize Design-Build-Test-Learn cycles, thereby transforming research methodologies. As illustrated in Figure 2, the hierarchical structure of robotics and AI integration in autonomous laboratories underscores the pivotal roles of SDLs and the synergy between robotics and AI. This figure categorizes key functions, technological integrations, challenges, and innovations, providing a comprehensive overview of their transformative impact on research methodologies and productivity in materials science. The following subsection delves

into SDLs' specific functions and implications, highlighting their role in the future of autonomous experimentation.

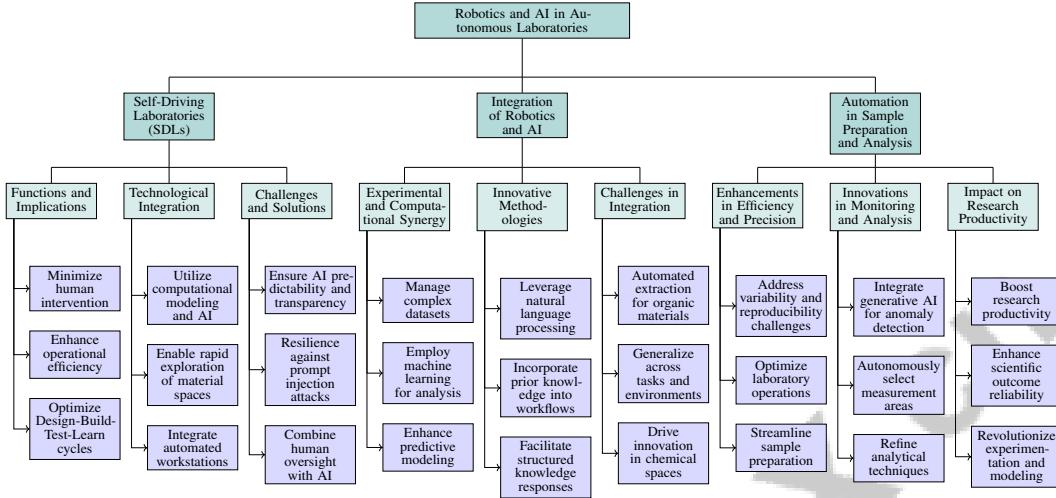


Figure 2: This figure illustrates the hierarchical structure of robotics and AI integration in autonomous laboratories, highlighting the roles of Self-Driving Laboratories (SDLs), the synergy between robotics and AI, and advancements in automation for sample preparation and analysis. Key functions, technological integrations, challenges, and innovations are categorized to provide a comprehensive overview of the transformative impact on research methodologies and productivity in materials science.

3.1 Role of Self-Driving Labs (SDLs)

Self-Driving Laboratories (SDLs) are pivotal in advancing autonomous experimentation by reducing human oversight and enhancing research efficiency through closed Design-Build-Test-Learn cycles [41]. Utilizing sophisticated computational modeling and AI, SDLs enable rapid exploration of material spaces [42]. MatPilot exemplifies this framework by integrating automated workstations for material preparation and testing, significantly accelerating research [43, 44]. AI-driven systems like CrystaLLM automate crucial tasks such as crystal structure prediction [45].

Advanced cognitive architectures enhance SDLs' long-term autonomy and deployment in complex environments [46]. CAMEO, a closed-loop autonomous system, optimizes materials exploration through iterative experimentation and learning [47]. Bayesian active learning in automated scanning probe microscopy (SPM) exemplifies how SDLs improve research efficiency [48].

Challenges remain in ensuring AI predictability, operational transparency, and user trust, particularly concerning AI errors or harmful actions [40]. Resilience against prompt injection attacks is crucial for SDL robustness [49]. A balanced approach, combining human oversight with AI, is essential for optimizing decision-making in automated experiments [50]. Embracing these innovations, SDLs are poised to revolutionize materials science research, accelerating new material discovery with enhanced speed and precision.

3.2 Integration of Robotics and AI for Experimental and Computational Synergy

The synergy between robotics and AI in materials science and chemistry marks a significant advancement, enhancing both experimental and computational research. This integration is crucial for managing complex datasets from large-scale experiments, requiring advanced machine learning techniques for effective analysis [51]. The LLMatDesign framework demonstrates this integration by employing LLMs for iterative material design, streamlining discovery workflows [1].

Innovative methodologies, such as ElaTBot, leverage natural language processing to predict elastic properties from material data, highlighting AI's potential in predictive modeling [52]. The

SEEK method incorporates prior knowledge into deep kernel learning-driven microscopy workflows, enhancing exploration efficiency [53].

LLMs like Polymetis facilitate structured knowledge responses in materials science, aiding both experimental and computational processes [9]. Cephalo V-LLMs enhance this integration by combining visual and linguistic data for improved interpretation [2]. Gaussian Process Regression (GPR), with advanced techniques, improves the synergy between experimental and computational processes by accurately modeling outcomes [6]. Automated laboratories, integrating robotics and AI, promise significant advancements by applying machine learning in experimental settings [7].

Challenges persist, particularly in automated extraction methods for organic materials, which lag behind those for inorganic materials [8]. Addressing these challenges is crucial for advancing robotics and AI's role, paving the way for more intelligent and adaptive systems. Embodied agents' architecture must generalize across tasks and environments to facilitate robust learning [54]. The integration of robotics and AI within materials science and chemistry not only enhances research processes but also drives innovation by enabling precise exploration of vast chemical spaces.

3.3 Automation in Sample Preparation and Analysis

Automation in sample preparation and analysis is fundamental to modern materials science, greatly enhancing experimental workflow efficiency and precision. As illustrated in Figure 3, the hierarchical structure of automation encompasses key components such as Materials Acceleration Platforms, Autonomous Monitoring, and Sample Analysis Automation. These elements work together to enhance efficiency and precision in materials science through advanced robotics and AI integration. Materials acceleration platforms (MAPs) have been crucial in addressing variability and reproducibility challenges of manual experimentation [27]. These platforms utilize advanced robotics and AI to automate complex tasks, reducing human intervention and error rates.

AlabOS, a Python-based reconfigurable workflow system, exemplifies advancements in this domain by employing a graph-based model for real-time sample tracking and resource allocation, optimizing laboratory operations [42]. Such innovations highlight automation's potential to streamline sample preparation, enhancing material characterization speed and accuracy.

In autonomous monitoring, integrating generative AI with robotics offers a novel approach to environmental anomaly detection. Quadruped robots equipped with this technology autonomously navigate labs, identifying deviations from expected conditions [55]. This capability is vital for maintaining consistent experimental conditions and preventing data discrepancies.

The AERM method, using active learning and Gaussian process regression, exemplifies automation in sample analysis by autonomously selecting measurement areas based on predicted uncertainties, improving data acquisition efficiency [56]. Automation's transformative impact is evident in refining analytical techniques and accelerating novel materials discovery.

These advancements in automation for sample preparation and analysis demonstrate profound improvements in efficiency and precision through robotics and AI integration. Overcoming manual process limitations, automation significantly boosts research productivity and enhances scientific outcome reliability and reproducibility. This transformation is especially notable in chemical sciences, where hardware automation and machine learning have revolutionized experimentation and modeling. Automated robotic systems improve experimental efficiency and accuracy, while AI-driven frameworks streamline data collection and analysis, allowing researchers to discern patterns and optimize synthesis processes. Such innovations simplify automation for a broader range of researchers and address data management challenges, paving the way for robust scientific discoveries [44, 11, 15, 14].

4 Machine Learning and Deep Learning for Materials Discovery

The integration of machine learning (ML) and deep learning (DL) in materials science represents a transformative shift, enhancing the discovery and characterization of new materials. This section explores the foundational aspects of ML and DL, highlighting their role in modeling complex material behaviors and predicting properties with unprecedented accuracy. By examining these technologies' methodologies and applications, we gain insights into their capacity to address critical challenges within the field. The subsequent subsection delves into the specific contributions of ML and DL in

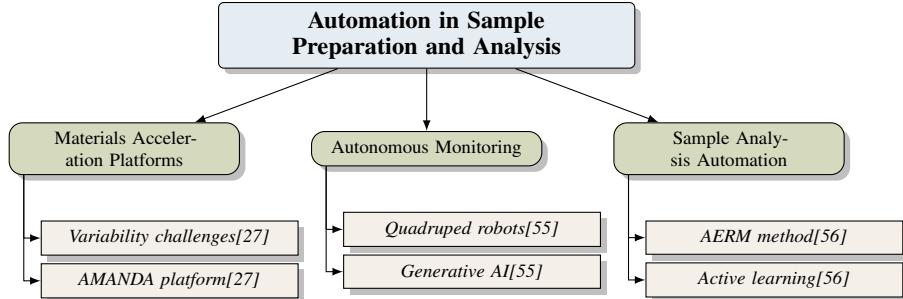


Figure 3: This figure illustrates the hierarchical structure of automation in sample preparation and analysis, highlighting key components such as Materials Acceleration Platforms, Autonomous Monitoring, and Sample Analysis Automation, which enhance efficiency and precision in materials science through advanced robotics and AI integration.

advancing materials discovery, paving the way for a deeper analysis of their innovative frameworks and techniques.

4.1 Machine Learning and Deep Learning

Machine learning (ML) and deep learning (DL) are crucial in advancing materials science by predicting material properties and exploring vast chemical spaces with precision and efficiency. These technologies facilitate sophisticated analysis of complex datasets, accelerating the discovery and development of novel materials. The AL-GAP method exemplifies active learning in selecting optimal configurations for fitting interatomic potentials, enhancing prediction accuracy from ab initio molecular dynamics [57].

The synergy between high-throughput calculations and ML models has been pivotal in predicting bulk materials' stability and properties, addressing core challenges in materials discovery [58]. DL's potential to transform materials discovery is underscored by its efficacy in property prediction and material synthesis, as evidenced by advancements in crystalline materials [4]. The AtomSets framework showcases ML models' adaptability in predicting material properties across diverse datasets [59].

Challenges persist, particularly in applying ML techniques to diverse and complex datasets, especially for modeling catalysts and their interactions [60]. Accurate property prediction, such as adsorbate and formation energies from atomic structure data, remains critical for efficient catalyst design [61]. The development of accurate and transferable ML potentials for predicting molecular energetics continues to require ongoing refinement [62].

ML frameworks utilizing validated datasets for conductivity and band gaps exemplify applications in predicting new transparent conducting materials (TCMs) [5]. Advanced techniques like extending Gaussian process regression models to account for non-i.i.d. noise and anisotropic behavior have enhanced ML's predictive capabilities in materials science [6].

However, current ML approaches face limitations in dynamic, non-stationary environments, particularly in robotics, highlighting the need for more robust and adaptable models [54]. The continued integration and refinement of ML and DL techniques promise significant advancements in materials science, driving innovation and enabling the discovery of new materials with enhanced properties and applications. By addressing existing challenges and harnessing these technologies' full potential, researchers can unlock new opportunities for exploration and development within the vast chemical landscape.

4.2 Innovative Machine Learning Frameworks

Innovative machine learning frameworks are revolutionizing materials discovery by integrating advanced computational techniques with experimental feedback, enabling efficient navigation of complex compositional spaces. Batch learning in high-throughput workflows exemplifies this integration,

enhancing input parameter selection and increasing density functional theory (DFT) calculations' success rates [63].

A modular approach to ML model building, incorporating active learning and descriptor selection, marks a significant advancement over previous benchmarks [61]. This approach facilitates adaptable models capable of efficiently handling diverse datasets, enhancing ML's robustness and applicability in materials research.

Hierarchical transfer learning frameworks, like AtomSets, demonstrate learning transfer from structural to compositional models, significantly enhancing prediction accuracy across material compositions [59]. This innovation highlights leveraging hierarchical structures to improve predictive models' adaptability and precision.

Active learning techniques optimize ML potentials by automatically selecting training data that enhances model performance while minimizing data requirements [62]. This strategy reduces the computational burden associated with large-scale materials simulations, accelerating the discovery process.

The Open MatSci ML Toolkit provides a flexible platform for implementing and experimenting with DL models, promoting novel ML methodologies exploration in materials science [60]. This toolkit exemplifies efforts to democratize access to advanced computational tools, fostering innovation and experimentation.

Developing learning systems that adapt to changing tasks and environments represents a fundamental improvement over traditional methods [54]. By focusing on adaptability and robustness, these systems are better equipped to handle materials research's dynamic nature, paving the way for more intelligent and responsive ML frameworks.

These innovative frameworks collectively highlight ML's transformative potential in materials discovery, leveraging advanced techniques like natural language processing (NLP) and DL. These approaches enable efficient navigation and insight extraction from vast chemical spaces, including automated literature classification, physical property extraction, and synthesis process identification. For instance, NLP pipelines categorize abstracts and extract key entities, while DL models demonstrate high efficacy in knowledge extraction from extensive datasets. This integration of ML methodologies enhances data retrieval precision and accelerates composition-structure-processing-property relationships exploration, advancing materials science significantly [16, 24, 18, 13].

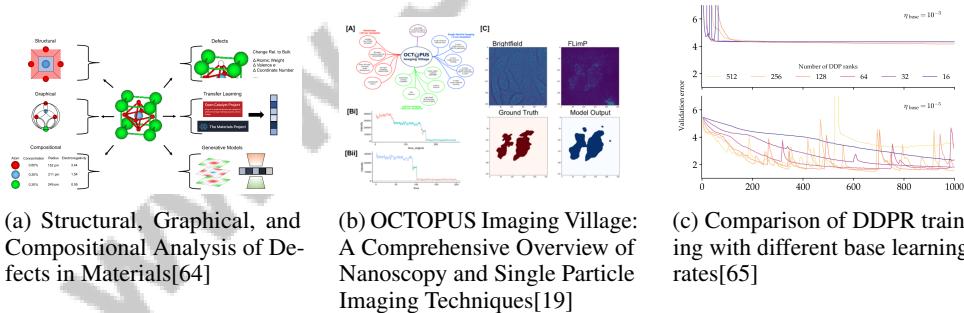


Figure 4: Examples of Innovative Machine Learning Frameworks

As shown in Figure 4, in materials discovery, integrating machine learning and deep learning frameworks has opened new research and development avenues. The examples in Figure 4 illustrate cutting-edge applications of these technologies. The first example, "Structural, Graphical, and Compositional Analysis of Defects in Materials," presents a flowchart outlining a sophisticated process for analyzing material defects, demonstrating how structural, graphical, and compositional data can be leveraged through transfer learning to generate novel materials. The second example, "OCTOPUS Imaging Village: A Comprehensive Overview of Nanoscopy and Single Particle Imaging Techniques," showcases a mind map categorizing various advanced imaging techniques, highlighting the central role of 'OCTOPUS Imaging Village' in connecting diverse methodologies such as Cryo-CLSM. The final example, "Comparison of DDPR training with different base learning rates," provides a comparative analysis of DDPR training methods, emphasizing varying base learning rates' impact on validation error across different DDPR ranks. Collectively, these examples underscore ML frame-

works' transformative potential in enhancing materials science research, facilitating breakthroughs in defect analysis, imaging techniques, and algorithmic training methodologies [64, 19, 65].

4.3 Deep Learning Techniques in Materials Science

Deep learning techniques have become central to advancing materials science research, offering unprecedented capabilities in data analysis, property prediction, and materials discovery. These techniques excel in managing complex datasets and identifying intricate patterns, facilitating extensive chemical space exploration. The MolTailor framework, which dynamically adjusts feature importance based on task descriptions, exemplifies innovation in molecular representation learning by integrating language models with molecular pretraining models [66].

The development of a dual-cutoff machine-learned potential (MLP) architecture represents a significant leap forward, enabling accurate modeling of short-range and long-range interactions, crucial for improving predictive capabilities in complex material systems [67].

In atomic force microscopy (AFM), domain transforms and unsupervised learning methods streamline atomic structure analysis, reducing manual labeling reliance and enhancing data processing workflows [68]. The SkipAtom model exemplifies DL's power by assuming atoms in similar chemical environments exhibit similar representations, facilitating accurate material property predictions [69].

Quantum active learning (QAL) methods further propel the field by employing quantum-enhanced feature maps and kernels to capture complex data relationships, increasing the likelihood of discovering optimal material properties [70]. This approach is particularly effective in structural determination tasks, where traditional methods may fall short.

The Materials Expert Artificial Intelligence (ME-AI) method represents a breakthrough in independently reproducing and expanding upon expert intuition, revealing new descriptors for predicting topological semimetals based on structural and chemical features [71]. This innovation underscores DL's potential to replicate and enhance human expertise in materials science.

Frameworks like AtomSets, which combine compositional and structural data using transfer learned features from graph networks, achieve state-of-the-art accuracy, demonstrating hierarchical transfer learning's effectiveness in materials research [59]. The Open MatSci ML Toolkit simplifies DL model development by integrating PyTorch Lightning for scalable computation and DGL for efficient graph neural network development, enhancing user experience [60].

Overall, DL techniques are transforming materials science by enabling efficient chemical space exploration, optimizing experimental and computational workflows, and facilitating novel materials discovery with tailored properties. As NLP and ML techniques advance, they are poised to significantly enhance knowledge extraction from the vast body of materials science literature, facilitating critical material property and synthesis process identification. This evolution is expected to unlock new innovation and discovery avenues, enabling researchers to navigate and leverage the wealth of experimental data efficiently, accelerating novel materials development and advancing the field [13, 16, 11, 18].

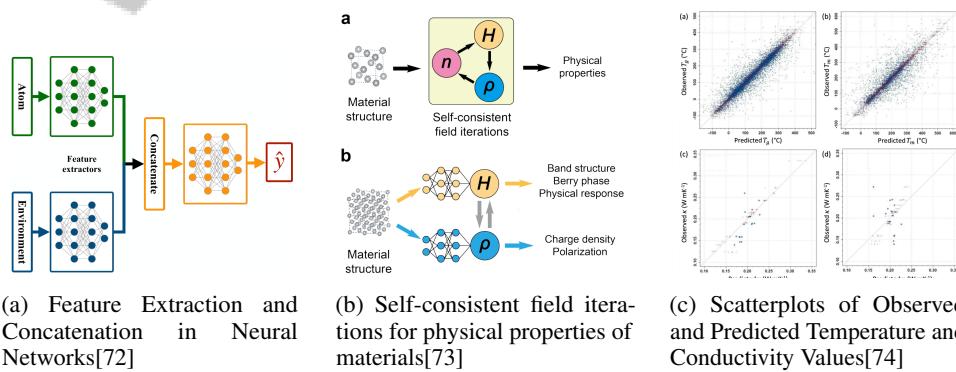


Figure 5: Examples of Deep Learning Techniques in Materials Science

As shown in Figure 5, the integration of machine learning and deep learning techniques in materials science has opened new pathways for materials discovery, offering innovative solutions for predicting and analyzing material properties. The examples illustrate diverse deep learning applications in this field. The first example, "Feature Extraction and Concatenation in Neural Networks," showcases a flowchart delineating the process of extracting and concatenating features from atoms and their environments using neural networks, highlighting the intricate information flow essential for material analysis. In the second example, "Self-consistent field iterations for physical properties of materials," the focus shifts to a diagrammatic representation of material structures and the iterative processes used to predict their physical properties, underscoring self-consistency's significance in computational modeling. Lastly, the "Scatterplots of Observed and Predicted Temperature and Conductivity Values" provide a visual comparison between observed and predicted data points, using scatterplots to evaluate temperature and conductivity prediction accuracy. These examples collectively demonstrate how deep learning techniques are pivotal in advancing materials science by enhancing materials discovery and analysis precision and efficiency [72, 73, 74].

4.4 Large Language Models and Automated Knowledge Extraction

Large Language Models (LLMs) have become pivotal tools in automated knowledge extraction, profoundly impacting materials science by enhancing data interpretation and facilitating novel materials discovery. The LLMatDesign framework exemplifies LLMs' transformative potential, operating in a zero-shot manner to swiftly adapt to new tasks through natural language prompts, streamlining material discovery processes [1]. This capability underscores LLMs' efficiency in navigating complex chemical spaces and improving decision-making accuracy.

ElaTBot, another innovative application, showcases LLMs' superiority over traditional ML approaches in predicting elastic constant tensors, highlighting their potential in advancing materials science applications [52]. Similarly, Polymetis demonstrates LLMs' ability to provide accurate and organized knowledge responses, emphasizing their role in enhancing research methodologies [9].

Integrating LLMs into symbolic regression tasks, as demonstrated by Hu et al., leverages depth-first search and reflection mechanisms to optimize formula generation, improving materials law discovery accuracy and efficiency [75]. This integration is crucial for advancing predictive modeling and understanding complex material behaviors.

The Materials Expert Artificial Intelligence (ME-AI) framework utilizes Gaussian process regression to discover emergent descriptors from expertly curated datasets, facilitating topological semimetals prediction [71]. This approach illustrates LLMs' potential to replicate and enhance expert intuition in materials science.

The systematic evaluation of multimodal language models by MaCBench highlights understanding the interplay between different modalities across scientific workflows, rather than focusing on isolated tasks [38]. This comprehensive assessment is vital for optimizing LLMs' integration in materials research.

Optimizing parameters at training time using LLMs, as discussed by Lin et al., enhances AI systems' safety and process supervision, further solidifying LLMs' role in improving research outcomes [39]. Collectively, these advancements illustrate LLMs' profound impact in revolutionizing materials science by enhancing predictive capabilities, improving data-driven insights, and accelerating new materials discovery. As the field evolves, leveraging LLMs' full potential will be instrumental in achieving significant breakthroughs in materials discovery and development.

5 High-throughput Screening and DFT in Computational Chemistry

To understand the progress in high-throughput screening (HTS) within computational chemistry, it is essential to explore the foundational principles and techniques underpinning this approach. The following subsection delves into specific HTS techniques crucial for accelerating material discovery, highlighting their role in overcoming the limitations of traditional density functional theory (DFT) simulations.

5.1 High-throughput Screening Techniques

High-throughput screening (HTS) techniques are pivotal in materials science, significantly expediting the discovery and development of novel materials. These methodologies enable systematic exploration of extensive parameter spaces, effectively mitigating the computational costs and time constraints inherent in traditional DFT simulations [76]. The integration of HTS with machine learning models further enhances predictive accuracy, as demonstrated by efficient voltage predictions for electrode materials [77].

The MODNet method exemplifies HTS's power by screening over 30 million structures from federated databases to identify high-refractive-index materials, showcasing HTS's capability to navigate complex material landscapes [78]. Automated workflows expedite the exploration of metastable phases, leveraging large-scale datasets, including DFT calculations of 20 million off-equilibrium conformations of over 50,000 organic molecules [51, 79].

In organic compounds, HTS combined with diffusion-based generative design methods has successfully identified new materials with favorable formation energies, underscoring HTS's transformative impact [80]. Additionally, datasets of thousands of atomic structures representing diverse adsorbate-surface interactions are critical for enhancing machine learning applications in this domain [61].

HTS techniques represent a transformative approach in materials science, advancing the speed and accuracy of material discovery. By utilizing advanced automation and large-scale data analysis, HTS methodologies facilitate significant progress in identifying and developing novel material candidates. Recent studies have demonstrated the effectiveness of natural language processing (NLP) tools in extracting insights from extensive materials science literature, enabling efficient analysis of physical properties, synthesis processes, and material usage trends [13, 24, 18, 16].

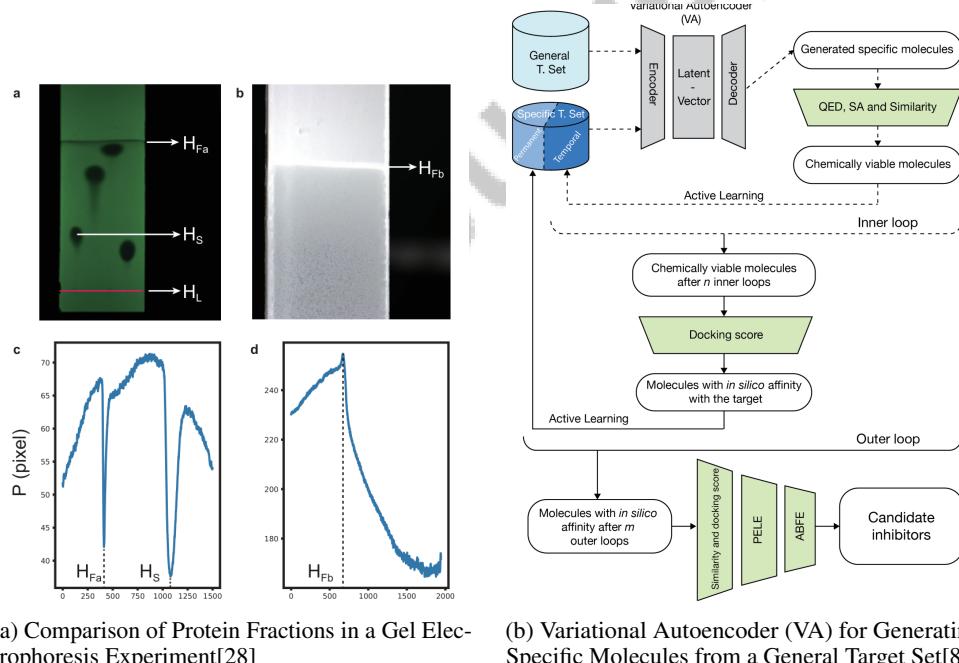


Figure 6: Examples of High-throughput Screening Techniques

As illustrated in Figure 6, HTS techniques are crucial in computational chemistry, providing a robust platform for rapid assessment and identification of potential compounds and materials. The synergy between HTS and DFT enhances exploration of chemical spaces. Two illustrative examples include the comparative analysis of protein fractions in gel electrophoresis and the use of a Variational Autoencoder (VA) for molecule generation. The former employs gel electrophoresis images and statistical tools like histograms and cumulative distribution functions (CDFs) to elucidate protein separation based on molecular weights. The latter demonstrates a sophisticated molecular generation approach using a VA, emphasizing machine learning's potential in chemical synthesis. Together, these

examples underscore HTS techniques' versatility and power in advancing computational chemistry research [28, 81].

5.2 Integration of DFT and Machine Learning

The integration of Density Functional Theory (DFT) and machine learning (ML) marks a significant advancement in computational chemistry, combining quantum mechanical precision with computational efficiency. This synergy enhances exploration of vast chemical compound spaces (CCS) by leveraging DFT's accuracy in electronic structure calculations alongside the predictive power of ML models [82]. A primary challenge in employing DFT is its high computational cost, which limits the scalability necessary for comprehensive materials analysis [83]. Innovative approaches such as the FDLS training method combine DFT and experimental data to create more accurate ML potentials for materials [84].

The SchNOrb framework exemplifies ML integration with quantum chemistry methods, enhancing electronic structure predictions through its synergy with DFT [85]. Similarly, the DC-MLP employs a dual-cutoff strategy to utilize short-range and long-range descriptors for intra- and inter-molecular interactions, improving ML model accuracy in predicting material properties [67].

The DFT IntML approach optimizes input parameters for DFT calculations through iterative learning from prior computational results, showcasing ML's potential to refine methodologies and improve predictive accuracy [63]. Furthermore, AI screening integrated with DFT calculations enables high-accuracy filtering of material candidates, enhancing computational chemistry efficiency [82].

Effective atomic and compound representations can be derived from chemical composition alone, as demonstrated by the SkipAtom model, which performs competitively in predicting material properties without structural information [69]. This capability is crucial for advancing predictive modeling and understanding complex material behaviors.

The integration of DFT and ML significantly enhances methodologies for materials discovery and characterization, facilitating accelerated simulations of electronic structures while reducing computational resource demands. By optimizing DFT calculations with ML techniques, researchers can explore larger, more complex systems, identifying new materials and chemical reaction pathways. Incorporating ML into atomistic modeling and high-dimensional interatomic potentials reshapes workflows in structure prediction, dynamics, and spectroscopy, streamlining computational processes and fostering a deeper understanding of underlying physics [86, 16, 87, 88, 74]. The combination of DFT's strengths with ML's predictive capabilities fosters more efficient and accurate computational chemistry methods, driving innovation and facilitating the discovery of novel materials with tailored properties.

5.3 Challenges in DFT Calculations

The application of Density Functional Theory (DFT) in computational chemistry faces several significant challenges that limit its efficacy in large-scale materials discovery. A primary obstacle is the computational intensity of DFT, which scales polynomially with system size, restricting its applicability to larger molecular systems [87]. This computational burden is compounded by the resource-intensive nature of traditional quantum chemical techniques, making exhaustive assessments of numerous potential layered structures infeasible [89].

Another critical challenge is the scarcity of high-quality training data essential for developing accurate ML models that complement DFT calculations [90]. The lack of robust datasets limits the ability to effectively train models capable of predicting material properties with the required precision for meaningful materials discovery.

Moreover, accurately modeling interactions within chemical compound spaces (CCS) presents a complex challenge due to the intricate electronic interactions DFT aims to capture [90]. This complexity is exacerbated by inefficiencies associated with current methods that fail to effectively represent the density matrix, leading to further computational inefficiencies in DFT calculations [73].

To address the computational challenges associated with DFT, it is essential to explore innovative strategies that alleviate the computational burden while ensuring accuracy in simulations. Recent advancements in machine learning and natural language processing offer promising avenues for

enhancing DFT calculation efficiency, enabling larger-scale simulations and extracting valuable insights from scientific literature. These approaches facilitate automated knowledge extraction, streamline data analysis, and accelerate materials discovery and the investigation of complex chemical reaction pathways [16, 88, 13, 91, 18]. Developing ML techniques to predict electronic properties with reduced reliance on exhaustive DFT calculations is vital for accelerating the discovery process. Additionally, improving density matrix representation and enhancing training data quality and availability are crucial for overcoming traditional DFT limitations. Addressing these challenges can enhance DFT scalability and applicability in computational chemistry, paving the way for comprehensive exploration of chemical spaces.

5.4 Advancements in DFT-based Methods

Recent advancements in DFT-based methods have significantly impacted computational chemistry by addressing traditional approaches' limitations and enhancing materials discovery efficiency. Machine learning (ML) has played a pivotal role in this transformation, offering innovative solutions to overcome computational constraints inherent in conventional DFT calculations [88]. Integrating ML algorithms with DFT has led to methodologies that accelerate high-throughput searches for novel materials, achieving remarkable speed improvements [92].

Key advancements include algorithms capable of constructing convex hulls for alloy systems, leading to new stable structure discoveries. These algorithms demonstrate speedups of three to four orders of magnitude compared to traditional high-throughput DFT calculations, significantly enhancing materials screening efficiency [92]. This improvement is crucial for exploring vast chemical compound spaces and identifying promising candidates for further investigation.

Moreover, integrating ML techniques with DFT enables greater accuracy and reduced computational costs in predicting material properties. These advancements facilitate exploring complex material systems and identifying novel compounds with tailored properties, driving innovation in computational chemistry [88]. The ability to efficiently predict electronic structures and material stability has profound implications for developing new materials and advancing various technological applications.

Recent advancements in DFT-based methods highlight the transformative potential of integrating traditional quantum mechanical techniques with modern machine learning approaches. By leveraging DFT's strengths for accurate predictions of molecular energies and electron densities and combining them with ML models that offer significant computational efficiency, researchers are developing innovative frameworks that reduce resource requirements and enhance electronic structure simulations' accuracy. This synergy allows for modeling larger molecular systems and facilitates in-silico materials discovery and chemical reaction pathway exploration, pushing the boundaries of computational chemistry and materials science [87, 88, 90]. By leveraging these innovations, researchers can overcome conventional DFT methodologies' limitations, paving the way for more efficient and comprehensive exploration of chemical spaces in computational chemistry.

6 Active Learning and Automated Experimentation

Category	Feature	Method
Integration of Active Learning in Experimental Workflows	Uncertainty-Based Selection	AL-QBC[62]
Automated Experimentation Techniques	Predictive Adjustment Techniques Efficiency Enhancement Methods Autonomous Process Management	GEM[93] AERM[56], HL[94], AL-NNT[95] pyron[33], AiIDA-d[96], N/A[97]
Real-time Monitoring and Anomaly Detection	Generative AI Integration	QRGAM[55]

Table 1: This table presents a comprehensive overview of various methodologies applied in the integration of active learning within experimental workflows, automated experimentation techniques, and real-time monitoring and anomaly detection in materials science. It categorizes the methods based on their primary focus areas, including uncertainty-based selection, predictive adjustment techniques, efficiency enhancement methods, and generative AI integration. The table highlights key features and specific methods, citing relevant studies that demonstrate their application and impact on advancing research efficiency and precision.

In materials science, innovative methodologies are crucial for boosting research efficiency and effectiveness. Active learning techniques, which optimize experimental workflows by strategically

selecting data points for analysis, are central to this evolution. This approach not only streamlines experimentation but also integrates advanced computational methods into traditional research frameworks. Table 2 provides a detailed overview of the methodologies employed in active learning, automated experimentation, and real-time monitoring within materials science, highlighting their contributions to enhancing research efficiency and accuracy. The following subsection explores the integration of active learning in experimental workflows, emphasizing its transformative impact on materials discovery and development.

6.1 Integration of Active Learning in Experimental Workflows

Active learning significantly optimizes experimental workflows by improving materials discovery efficiency and reducing exhaustive testing requirements. This methodology employs machine learning models to identify the most informative data points, refining datasets and enhancing predictive accuracy [62]. By automating dataset generation, active learning minimizes redundant calculations and human intervention, optimizing resource allocation.

The combination of active learning with Gaussian Process Regression (GPR) exemplifies its potential to predict material properties with fewer measurements than traditional methods, thereby enhancing experimental efficiency [62]. This synergy enables a more focused exploration of chemical spaces, ensuring computational and experimental resources target the most promising candidates.

Methodologies such as the hypothesis learning method by Ziatdinov et al. leverage structured Gaussian Processes and reinforcement learning for effective navigation of experimental and hypothesis spaces. This approach enables exploration of complex parameter spaces with minimal steps, as demonstrated in studies of concentration-induced phase transitions in combinatorial materials libraries [94, 98, 18, 99, 47]. By integrating Bayesian inference with structured Gaussian Processes, this method reduces uncertainty in material property exploration, facilitating informed decision-making.

In high-throughput searches, active learning algorithms dynamically generate and refine training datasets, approximating quantum-mechanical energies with machine-learning interatomic potentials [62]. This dynamic refinement accelerates discovery by focusing computational efforts on the most informative regions of chemical space, thereby optimizing experimental workflows.

The integration of active learning in experimental workflows marks a significant advancement in materials science, enabling researchers to explore vast chemical spaces with greater precision and speed. By concentrating on the most informative data points and refining methodologies, active learning enhances experimental efficiency and fosters innovation in materials discovery. This approach leverages FAIR (findable, accessible, interoperable, and reusable) data management principles to optimize workflows, with studies demonstrating a tenfold reduction in simulations required to identify optimal alloy compositions. Additionally, active learning techniques, including closed-loop autonomous systems, facilitate real-time hypothesis testing and rapid material optimization, allowing researchers to navigate complex synthesis-process-structure-property landscapes more effectively [100, 99, 34, 47].

6.2 Automated Experimentation Techniques

Automated experimentation techniques are transforming materials science research by enhancing efficiency, reproducibility, and data quality. These techniques employ advanced robotics and machine learning to automate complex experimental tasks, reducing human intervention and improving research consistency. The Clio robotic platform exemplifies these advancements by autonomously dosing, mixing, and measuring various electrolyte formulations, showcasing automation's potential in optimizing experimental processes [97].

Integrating active learning strategies within automated systems further refines experimental workflows by dynamically selecting the most informative data points for analysis. For instance, the AL-NNT method combines Monte Carlo sampling with artificial neural network (ANN) training to effectively utilize sparse datasets, creating robust models that optimize experimental methodologies [95]. Similarly, the Gemini method employs neural networks for dynamic bias correction, modeling relationships between inexpensive and costly measurements to inform decision-making in optimization tasks, enhancing precision and efficiency [93].

Automated systems like AiiDA-defects illustrate the benefits of automation in complex computational tasks, such as defect calculations, by alleviating scientists' workloads and improving result reliability [96]. These systems promote collaboration among researchers by streamlining workflows and ensuring reproducibility, critical for advancing scientific research [33].

Moreover, the hypothesis learning method, employing a warm-up followed by an exploration phase, showcases the efficiency of automated systems in navigating parameter spaces and refining hypotheses [94]. The autonomous selection and measurement of specific areas within a materials library, as demonstrated by Thelen et al., further optimize characterization by focusing on regions with the highest predicted uncertainty, enhancing data acquisition precision and efficiency [56].

Future work in automated experimentation should prioritize developing better evaluation metrics and robust learning systems capable of operating safely in diverse environments, addressing scalability and adaptability challenges in various research settings [54]. Automated experimentation techniques are transforming research methodologies by enhancing reproducibility, reducing technical complexity, and fostering collaboration among researchers. These advancements streamline experimental workflows and pave the way for more efficient and reliable scientific outcomes, driving innovation and discovery in materials science.

6.3 Real-time Monitoring and Anomaly Detection

Real-time monitoring and anomaly detection are crucial for enhancing the accuracy and reliability of experimental processes in materials science. These methodologies utilize advanced technologies to continuously observe experimental conditions and promptly identify deviations from expected outcomes, ensuring data integrity and precision. Integrating real-time monitoring systems with automated experimentation frameworks allows researchers to maintain stringent control over experimental variables, significantly reducing the likelihood of errors or inconsistencies [55].

An exemplary application of real-time monitoring is the use of generative AI integrated with robotics to autonomously navigate laboratory environments. This approach enables the identification of environmental anomalies, ensuring that experimental setups remain within defined parameters and that any deviations are promptly addressed. Such capabilities are crucial for maintaining consistency and reliability in experimental data, particularly in complex research environments with limited manual oversight [55].

Furthermore, implementing advanced anomaly detection algorithms within autonomous systems enhances the ability to discern subtle deviations in experimental outputs that may indicate underlying issues. By employing machine learning techniques, these systems can learn from historical data to predict potential anomalies, allowing for proactive adjustments to experimental conditions. This predictive capability improves experimental accuracy by enabling targeted data collection and informed decision-making while optimizing resource allocation by minimizing redundant testing and corrections, thus facilitating a more efficient materials discovery process [13, 24, 101, 16].

The integration of real-time monitoring and anomaly detection within autonomous laboratories exemplifies the transformative potential of these technologies in advancing materials science research. By conducting experiments under optimal conditions and promptly addressing deviations, these methodologies enhance the reliability and reproducibility of scientific outcomes, as evidenced by thorough workflows and well-documented codebases in pioneering studies. Such practices facilitate successful result replication and set commendable standards for future research. Moreover, analyzing machine learning techniques in materials science reveals that addressing common pitfalls, such as imbalanced datasets and model introspection challenges, is crucial for achieving trustworthy predictions. Integrating explainable machine learning frameworks and promoting transparency in experimental processes contributes to more robust scientific inquiry and innovation across the field [16, 98, 102, 103, 13]. As the field evolves, developing sophisticated monitoring and detection systems will be essential for driving further innovations and discoveries in materials science.

Feature	Integration of Active Learning in Experimental Workflows	Automated Experimentation Techniques	Real-time Monitoring and Anomaly Detection
Optimization Technique	Gaussian Process Regression	Robotics And ML	Generative AI
Data Handling	Informative Data Selection	Sparse Dataset Utilization	Anomaly Prediction
Experimental Integration	Resource Allocation Optimization	Reproducibility Enhancement	Error Reduction

Table 2: This table presents a comparative analysis of methodologies employed in active learning, automated experimentation, and real-time monitoring within materials science. It highlights the optimization techniques, data handling strategies, and experimental integration processes associated with each approach. The table underscores the contributions of these methodologies to enhancing research efficiency and accuracy in the field.

7 Challenges and Future Directions

7.1 Challenges in Multi-Robot Coordination

Coordinating multiple robots in autonomous laboratories presents significant challenges due to the complexities of integrating diverse robotic capabilities into a cohesive operational framework. As robotic implementation scales, task planning and execution become more intricate, necessitating advanced algorithms for effective interaction management. A major challenge is the dynamic allocation of robotic resources, as device availability fluctuates during runtime, requiring real-time adjustments for optimal utilization [62].

Reliance on visual input and occasional inaccuracies in environmental perception can hinder task execution, emphasizing the need for robust perception systems to ensure reliable operations in unpredictable environments [54]. Additionally, the complexity of programming custom control software for scientific instruments restricts the accessibility and scalability of automated solutions.

A critical issue in autonomous laboratories is dependence on pre-synthesized samples, complicating coordination among robots exploring new materials. This is further exacerbated by the computational costs associated with machine learning interatomic potentials (MLIPs), especially in high-dimensional spaces where evaluating expected information gain can be resource-intensive [57]. Existing benchmarks inadequately assess MLIPs on disordered systems with multiple components, leading to gaps in understanding their accuracy and transferability.

The brittleness and limited versatility of robots further complicate these challenges, as they often fail to operate reliably outside well-defined conditions [54]. Addressing these challenges is vital for enhancing the capabilities of autonomous laboratory systems and improving the efficiency and reliability of multi-robot coordination in complex experimental environments.

7.2 Data Quality and Availability

Data quality and availability are pivotal challenges in integrating artificial intelligence (AI) and automation within materials science and chemistry, significantly impacting machine learning (ML) applications' effectiveness. The performance of ML models hinges on the quality of training data, as demonstrated by the LLMatDesign framework, where the quality of large language models (LLMs) influences the accuracy and reliability of suggested modifications [1]. Similarly, the Cephalo multimodal vision language models emphasize the necessity of high-quality datasets, affecting their ability to interpret material phenomena [2].

The scarcity of high-quality, standardized datasets poses a substantial barrier to robust ML model development, limiting predictive accuracy and applicability [3]. This challenge is compounded by biases introduced by initial training datasets, as seen in optical materials discovery, where promising candidates may be overlooked [78]. The lack of comprehensive experimental datasets accurately representing diverse chemical properties complicates data-driven predictions for band gaps [5].

Existing benchmarks often fail to evaluate the integration of multiple modalities, leading to a lack of understanding of models' abilities in complex reasoning and spatial understanding, crucial in scientific contexts [38]. This inadequacy highlights the need for refined benchmarks that capture materials science research nuances.

The complexity of chemical systems and the challenges of integrating ML with traditional chemical knowledge further exacerbate these issues, as current research is often hampered by limited access

to high-quality data [7]. Additionally, the sensitivity of LLMs to different prompts complicates the generation and refinement of parameters in complex systems [39].

Establishing standardized databases and enhancing data curation processes is essential to tackle these challenges effectively. This will improve data quality and availability and facilitate the extraction of structured, actionable insights from the vast amounts of unstructured information present in the literature. Advanced frameworks utilizing natural language processing and LLMs can streamline the categorization and retrieval of material properties and synthesis processes, accelerating materials discovery and innovation. Developing a robust data ecosystem that integrates ML models with diverse data sources will support ongoing research and enable efficient knowledge dissemination across the materials science community [12, 16, 13, 20]. Overcoming these limitations will unlock new opportunities for innovation and discovery, advancing the field through more effective and reliable AI-driven methodologies.

7.3 Model Interpretability and Accuracy

Model interpretability and accuracy are fundamental to AI-driven research in materials science and chemistry, significantly influencing the reliability and utility of AI predictions in guiding experimental and computational efforts. Interpretability is crucial for understanding the mechanisms behind AI predictions, enabling researchers to infer causation and assess uncertainty estimates in model explanations [104]. This understanding fosters trust among researchers, essential for integrating AI methodologies into scientific workflows [17].

The opacity of machine learning algorithms, particularly neural networks, poses a challenge by hindering their adoption and understanding among researchers in physical sciences [37]. This opacity complicates inferring causation, necessitating the development of models offering greater transparency and explanatory power. The need for interpretability is further emphasized in frameworks like H-CLMP, where predictions in new composition spaces require clear explanations to ensure confidence in results [105].

Accuracy is equally critical, underpinning the reliability of predictions guiding scientific exploration. Achieving high accuracy in energy and force predictions with reduced training datasets, as shown by certain active learning approaches, is valuable for high-fidelity quantum calculations [95]. However, challenges persist, such as the stability of large language models (LLMs) in continuous quantitative predictions, which can be affected by minor variations in conditions like temperature [52].

The integration of multimodal information in AI models presents additional challenges, as current VLLMs struggle with scientific reasoning and effectively integrating diverse data types, highlighting the need for improved model architectures [38]. Despite these challenges, leveraging both composition and structure information can lead to lower prediction errors and improved performance, as seen in multimodal machine learning approaches [106].

Enhancing model interpretability and accuracy is crucial for advancing AI-driven research in materials science and chemistry. By systematically addressing challenges associated with knowledge extraction from the expanding body of scientific literature, researchers can enhance the reliability and depth of their models, streamlining the research process and accelerating scientific discovery and innovation. Leveraging advanced techniques such as named-entity recognition (NER) and deep learning can facilitate the automatic extraction of critical insights from vast datasets, as demonstrated in recent studies focused on organic materials. These advancements will unlock the full potential of AI technologies across various scientific domains, enabling more effective exploration and understanding of complex phenomena [15, 18].

7.4 Integration of Computational and Experimental Methods

The integration of computational and experimental methods is essential for advancing research outcomes in materials science and chemistry, offering a comprehensive approach to exploring chemical spaces with increased accuracy and efficiency. This synergy leverages computational techniques, such as machine learning (ML) and high-throughput experimentation, to refine experimental workflows and enhance the predictive accuracy of material properties. The SchNOrb framework exemplifies this integration by unifying machine learning with quantum chemistry methods, improving electronic structure predictions and facilitating novel material discovery [85].

Future research should focus on refining active learning strategies and exploring their applicability to complex chemical systems beyond organic molecules, as suggested by Sivaraman et al., to enhance the training process and predictive power of machine learning force fields (MLFFs) [57]. Additionally, integrating physical principles into ML models and developing more intuitive methodologies, as highlighted by Sajjan et al., remain crucial for advancing quantum machine learning in chemistry [37].

The Botfip-LLM framework illustrates the potential of integrating computational methods with experimental approaches by enhancing the processing and understanding of symbolic formula information [107]. This integration is critical for advancing predictive modeling capabilities, particularly in automated characterization processes where ML techniques can optimize experimental workflows. Future research should also explore expanding optimization techniques to include multimodal systems and integrating architecture search methods to improve compound AI system performance [39].

Moreover, enhancing scalability and incorporating more complex 3D structures, as suggested by Jørgensen et al., will improve exploration efficiency in diverse chemical spaces [108]. Future research should also focus on expanding datasets and enhancing the integration capabilities of benchmarks with other data services, as noted by Blaiszik et al., to support ML applications [12].

The integration of theoretical databases with experimental data is essential for developing automated workflows that facilitate materials discovery. This approach enables researchers to explore new ML techniques enhancing predictive capabilities and driving innovation in materials science. Future research should focus on enhancing the toolkit's capabilities by integrating additional datasets and benchmarks and exploring new model architectures to advance materials discovery [60].

The integration of computational and experimental methods holds significant promise for advancing materials science, enabling researchers to explore complex chemical spaces with greater precision and efficiency. By systematically addressing current challenges in knowledge extraction and leveraging emerging technologies such as natural language processing (NLP) and large language models (LLMs), the materials science field can enhance its ability to discover and develop new materials. This includes automating the extraction of critical information from vast amounts of unstructured literature, streamlining the research process, and enabling researchers to identify promising experimental avenues more efficiently. Future research directions should focus on refining these computational methods and building high-quality, multi-modal datasets to support hypothesis generation and testing, ultimately driving significant breakthroughs in materials discovery [16, 24, 13, 25, 18].

7.5 Scalability and Computational Efficiency

Scalability and computational efficiency are critical challenges in integrating advanced technologies within autonomous laboratories, particularly as these systems expand to accommodate increasingly complex scientific tasks. The scalability of these technologies is often constrained by the limitations of current computational frameworks, which may not efficiently handle the vast data and processing demands of large-scale experiments [44]. This is particularly relevant in systems dominated by long-range interactions, where existing implementations may fall short in providing accurate and efficient solutions [109].

A significant challenge lies in the trial-and-error approach commonly employed to determine suitable starting parameters for computational models, such as those used in magnetic multilayer systems. This method lacks systematic efficiency and can be cumbersome given the intricate nature of these systems [63]. The reliance on heuristic techniques underscores the need for robust methodologies that can streamline parameter selection and enhance the overall efficiency of computational workflows.

Moreover, questions remain regarding the reliability of outputs generated by large language models (LLMs) in specific scientific contexts, raising concerns about the dependability of these models in delivering consistent and accurate results [44]. Addressing these issues is crucial to ensure that autonomous laboratories can scale effectively and maintain computational efficiency while integrating LLMs and other AI-driven technologies.

Future research should focus on developing scalable computational frameworks that can efficiently manage the increased data and processing requirements of autonomous laboratories. This encompasses investigating innovative algorithmic strategies and refining current methodologies aimed at significantly improving the scalability and computational efficiency of systems involved in knowl-

edge extraction from materials science literature, leveraging advanced techniques such as natural language processing (NLP) and deep learning for enhanced data analysis and information retrieval [16, 24, 13, 91, 18]. By addressing these challenges, researchers can ensure that autonomous laboratories remain at the forefront of scientific innovation, capable of tackling increasingly complex problems with precision and speed.

7.6 Robustness and Generalization of Machine Learning Models

The development of robust and generalizable machine learning (ML) models is crucial for advancing materials science, as these models must effectively predict properties across diverse chemical systems [3]. The complexity of ML models, compounded by the growing size of datasets, presents challenges in understanding the relationship between input features and predicted properties, complicating the interpretability and scalability of these models [110]. This necessitates creating models capable of generalizing across different materials and capturing intricate phenomena [58].

Active learning strategies, such as SelectAL, have demonstrated significant improvements over traditional methods by offering flexibility and adaptability, performing well in both low and high-budget scenarios [99]. However, limitations arise when the method's performance plateaus with the inclusion of more than three modalities, suggesting potential constraints in current implementations [111]. Moreover, the scalability of Gaussian Processes to very large datasets remains a challenge, necessitating alternative modeling approaches in future work [61].

The AtomSets method, while effective in predicting numerous properties, may struggle with datasets requiring extensive structural information, highlighting the need for models that can handle diverse data types and distributions [59]. Additionally, the reliance on the quality of convolutional neural network (CNN) architecture and the computational resources required for training may restrict scalability to larger systems [108].

A critical limitation of current methodologies is the increased computational cost associated with optimizing for anisotropic kernels, which may require more complex hyperparameter tuning [6]. This challenge underscores the importance of developing efficient algorithms that can maintain robustness and generalization without incurring prohibitive computational expenses.

8 Conclusion

The convergence of robotics, artificial intelligence (AI), and autonomous laboratories is revolutionizing materials science and chemistry by streamlining research methodologies and enhancing precision and scalability. Self-Driving Laboratories (SDLs) epitomize this advancement, optimizing experimental workflows and expediting the discovery of new materials through closed-loop systems. Frameworks such as LLMatDesign exemplify the potential of these technologies to surpass conventional methods, underscoring the critical role of machine learning (ML) in accelerating materials design and property prediction. Despite these advancements, improving the accuracy of computational methods remains essential to broaden their applicability.

The synergy between high-throughput experimental techniques and ML-assisted diagnostics has significantly increased throughput, demonstrating the transformative impact of these approaches in advancing materials science. The AtomSets framework, with its effective deep learning model, illustrates the ability to accurately learn material properties across varied data sizes. Moreover, the AI workflow proposed by Lai et al. marks substantial progress in optimizing catalyst synthesis, highlighting the importance of refining synthesis parameters.

Looking ahead, research will focus on expanding models like Polymetis to enhance reasoning capabilities in response to the evolving needs of materials science. The application of DFT IntML has notably improved the efficiency of high-throughput calculations, enhancing success rates while reducing computational time. Additionally, Cephalo's ability to analyze complex material phenomena showcases its potential to propel research in bio-inspired materials and high-performance material design. The ML framework by Ottomano et al. further underscores the transformative potential in discovering new transparent conducting materials (TCMs), highlighting the necessity for continuous research and the integration of comprehensive experimental data.

References

- [1] Shuyi Jia, Chao Zhang, and Victor Fung. Llmatdesign: Autonomous materials discovery with large language models. *arXiv preprint arXiv:2406.13163*, 2024.
- [2] Markus J. Buehler. Cephalo: Multi-modal vision-language models for bio-inspired materials analysis and design, 2024.
- [3] Rampi Ramprasad, Rohit Batra, Ghanshyam Pilania, Arun Mannodi-Kanakkithodi, and Chiho Kim. Machine learning and materials informatics: Recent applications and prospects, 2017.
- [4] Zhenzhong Wang, Haowei Hua, Wanyu Lin, Ming Yang, and Kay Chen Tan. Crystalline material discovery in the era of artificial intelligence. *arXiv preprint arXiv:2408.08044*, 2024.
- [5] Federico Ottomano, John Y. Goulermas, Vladimir Gusev, Rahul Savani, Michael W. Gaulois, Troy D. Manning, Hai Lin, Teresa P. Manzanera, Emmeline G. Poole, Matthew S. Dyer, John B. Claridge, Jon Alaria, Luke M. Daniels, Su Varma, David Rimmer, Kevin Sanderson, and Matthew J. Rosseinsky. Assessing data-driven predictions of band gap and electrical conductivity for transparent conducting materials, 2024.
- [6] Marcus M. Noack, Gregory S. Doerk, Ruipeng Li, Jason K. Streit, Richard A. Vaia, Kevin G. Yager, and Masafumi Fukuto. Autonomous materials discovery driven by gaussian process regression with inhomogeneous measurement noise and anisotropic kernels, 2020.
- [7] Austin Cheng, Cher Tian Ser, Marta Skreta, Andrés Guzmán-Cordero, Luca Thiede, Andreas Burger, Abdulrahman Aldossary, Shi Xuan Leong, Sergio Pablo-García, Felix Strieth-Kalthoff, and Alán Aspuru-Guzik. Spiers memorial lecture: How to do impactful research in artificial intelligence for chemistry and materials science, 2024.
- [8] Nung Siong Lai, Yi Shen Tew, Xialin Zhong, Jun Yin, Jiali Li, Binhang Yan, and Xiaonan Wang. An artificial intelligence (ai) workflow for catalyst design and optimization, 2024.
- [9] Chao Huang, Huichen Xiao, Chen Chen, Chunyan Chen, Yi Zhao, Shiyu Du, Yiming Zhang, He Sha, and Ruixin Gu. Polymetis:large language modeling for multiple material domains, 2024.
- [10] Ganesh Narasimha, Dejia Kong, Paras Regmi, Rongying Jin, Zheng Gai, Rama Vasudevan, and Maxim Ziatdinov. Multiscale structure-property discovery via active learning in scanning tunneling microscopy, 2024.
- [11] Connor W. Coley, Natalie S. Eyke, and Klavs F. Jensen. Autonomous discovery in the chemical sciences part i: Progress, 2020.
- [12] Ben Blaiszik, Logan Ward, Marcus Schwarting, Jonathon Gaff, Ryan Chard, Daniel Pike, Kyle Chard, and Ian Foster. A data ecosystem to support machine learning in materials science, 2019.
- [13] Fusataka Kuniyoshi, Jun Ozawa, and Makoto Miwa. Analyzing research trends in inorganic materials literature using nlp, 2021.
- [14] Kexin Chen, Hanqun Cao, Junyou Li, Yuyang Du, Menghao Guo, Xin Zeng, Lanqing Li, Jiezhong Qiu, Pheng Ann Heng, and Guangyong Chen. An autonomous large language model agent for chemical literature data mining, 2024.
- [15] Yongchao Lu, Hong Wang, Lanting Zhang, Ning Yu, Siqi Shi, and Hang Su. Unleashing the power of ai in science-key considerations for materials data preparation. *Scientific Data*, 11(1):1039, 2024.
- [16] Vineeth Venugopal, Sourav Sahoo, Mohd Zaki, Manish Agarwal, Nitya Nand Gosvami, and N. M. Anoop Krishnan. Looking through glass: Knowledge discovery from materials science literature using natural language processing, 2021.
- [17] Brian DeCost, Jason Hattrick-Simpers, Zachary Trautt, Aaron Kusne, Eva Campo, and Martin Green. Scientific ai in materials science: a path to a sustainable and scalable paradigm, 2020.

-
- [18] Xintong Zhao, Steven Lopez, Semion Saikin, Xiaohua Hu, and Jane Greenberg. Text to insight: Accelerating organic materials knowledge extraction via deep learning, 2021.
 - [19] Tony Hey, Keith Butler, Sam Jackson, and Jeyarajan Thiyyagalingam. Machine learning and big scientific data, 2019.
 - [20] Mara Schilling-Wilhelmi, Martíño Ríos-García, Sherjeel Shabih, María Victoria Gil, Santiago Miret, Christoph T. Koch, José A. Márquez, and Kevin Maik Jablonka. From text to insight: Large language models for materials science data extraction, 2024.
 - [21] Johannes Betz, Hongrui Zheng, Zirui Zang, Florian Sauerbeck, Krzysztof Walas, Velin Dimitrov, Madhur Behl, Rosa Zheng, Joydeep Biswas, Venkat Krovi, and Rahul Mangharam. Teaching autonomous systems hands-on: Leveraging modular small-scale hardware in the robotics classroom, 2022.
 - [22] James Prather, Paul Denny, Juho Leinonen, Brett A. Becker, Ibrahim Albluwi, Michelle Craig, Hieke Keuning, Natalie Kiesler, Tobias Kohn, Andrew Luxton-Reilly, Stephen MacNeil, Andrew Peterson, Raymond Pettit, Brent N. Reeves, and Jaromir Savelka. The robots are here: Navigating the generative ai revolution in computing education, 2023.
 - [23] Bernard J. Koch and David Peterson. From protoscience to epistemic monoculture: How benchmarking set the stage for the deep learning revolution, 2024.
 - [24] Luca Foppiano, Guillaume Lambard, Toshiyuki Amagasa, and Masashi Ishii. Mining experimental data from materials science literature with large language models: an evaluation study, 2024.
 - [25] Santiago Miret and N M Anoop Krishnan. Are llms ready for real-world materials discovery?, 2024.
 - [26] Shijing Sun, Noor T. P. Hartono, Zekun D. Ren, Felipe Oviedo, Antonio M. Buscemi, Mariya Layurova, De Xin Chen, Tofunmi Ogunfunmi, Janak Thapa, Savitha Ramasamy, Charles Settens, Brian L. DeCost, Aaron Gilad Kusne, Zhe Liu, Siyu I. P. Tian, I. Marius Peters, Juan-Pablo Correa-Baena, and Tonio Buonassisi. Accelerating photovoltaic materials development via high-throughput experiments and machine-learning-assisted diagnosis, 2018.
 - [27] Jerrit Wagner, Christian G. Berger, Xiaoyan Du, Tobias Stubhan, Jens A. Hauch, and Christoph J. Brabec. The evolution of materials acceleration platforms – towards the laboratory of the future with amanda, 2021.
 - [28] Hao Xu, Jinglong Lin, Qianyi Liu, Yuntian Chen, Jianning Zhang, Yang Yang, Michael C. Young, Yan Xu, Dongxiao Zhang, and Fanyang Mo. High-throughput discovery of chemical structure-polarity relationships combining automation and machine learning techniques, 2022.
 - [29] Haotong Liang, Chuangye Wang, Heshan Yu, Dylan Kirsch, Rohit Pant, Austin McDannald, A. Gilad Kusne, Ji-Cheng Zhao, and Ichiro Takeuchi. Real-time experiment-theory closed-loop interaction for autonomous materials science, 2024.
 - [30] Cormac Toher, Corey Osse, David Hicks, Eric Gossett, Frisco Rose, Pinku Nath, Demet Usanmaz, Denise C. Ford, Eric Perim, Camilo E. Calderon, Jose J. Plata, Yoav Lederer, Michal Jahnátek, Wahyu Setyawan, Shidong Wang, Junkai Xue, Kevin Rasch, Roman V. Chepulkov, Richard H. Taylor, Geena Gomez, Harvey Shi, Andrew R. Supka, Rabih Al Rahal Al Orabi, Priya Gopal, Frank T. Cerasoli, Laalitha Liyanage, Haihang Wang, Ilaria Siloi, Luis A. Agapito, Chandramouli Nyshadham, Gus L. W Hart, Jesús Carrete, Fleur Legrain, Natalio Mingo, Eva Zurek, Olexandr Isayev, Alexander Tropsha, Stefano Sanvito, Robert M. Hanson, Ichiro Takeuchi, Michael J. Mehl, Aleksey N. Kolmogorov, Kesong Yang, Pino D'Amico, Arrigo Calzolari, Marcio Costa, Riccardo De Gennaro, Marco Buongiorno Nardelli, Marco Fornari, Ohad Levy, and Stefano Curtarolo. The aflow fleet for materials discovery, 2017.
 - [31] Sebastian Ament, Maximilian Amsler, Duncan R. Sutherland, Ming-Chiang Chang, Dan Guevarra, Aine B. Connolly, John M. Gregoire, Michael O. Thompson, Carla P. Gomes, and R. Bruce van Dover. Autonomous synthesis of metastable materials, 2021.

-
- [32] Seiji Takeda, Toshiyuki Hama, Hsiang-Han Hsu, Toshiyuki Yamane, Koji Masuda, Victoria A. Piunova, Dmitry Zubarev, Jed Pitera, Daniel P. Sanders, and Daiju Nakano. Ai-driven inverse design system for organic molecules, 2020.
 - [33] Sarath Menon, Yury Lysogorskiy, Alexander L. M. Knoll, Niklas Leimeroth, Marvin Poul, Minaam Qamar, Jan Janssen, Matous Mrovec, Jochen Rohrer, Karsten Albe, Jörg Behler, Ralf Drautz, and Jörg Neugebauer. From electrons to phase diagrams with classical and machine learning potentials: automated workflows for materials science with pyiron, 2024.
 - [34] Alex Wang, Haotong Liang, Austin McDannald, Ichiro Takeuchi, and A. Gilad Kusne. Benchmarking active learning strategies for materials optimization and discovery, 2022.
 - [35] Rachmad Vidya Wicaksana Putra, Alberto Marchisio, Fakhreddine Zayer, Jorge Dias, and Muhammad Shafique. Embodied neuromorphic artificial intelligence for robotics: Perspectives, challenges, and research development stack, 2024.
 - [36] Nikolaus Correll and Christoffer Heckman. Materials that make robots smart, 2017.
 - [37] Manas Sajjan, Junxu Li, Raja Selvarajan, Shree Hari Sureshbabu, Sumit Suresh Kale, Rishabh Gupta, Vinit Singh, and Sabre Kais. Quantum machine learning for chemistry and physics, 2022.
 - [38] Nawaf Alampara, Mara Schilling-Wilhelmi, Martíño Ríos-García, Indrajeet Mandal, Pranav Khetarpal, Hargun Singh Grover, N. M. Anoop Krishnan, and Kevin Maik Jablonka. Probing the limitations of multimodal language models for chemistry and materials research, 2024.
 - [39] Matthieu Lin, Jenny Sheng, Andrew Zhao, Shenzhi Wang, Yang Yue, Yiran Wu, Huan Liu, Jun Liu, Gao Huang, and Yong-Jin Liu. Llm-based optimization of compound ai systems: A survey, 2024.
 - [40] Sivan Schwartz, Avi Yaeli, and Segev Shlomov. Enhancing trust in llm-based ai automation agents: New considerations and future challenges, 2023.
 - [41] Hector Garcia Martin, Tijana Radivojevic, Jeremy Zucker, Kristofer Bouchard, Jess Sustarich, Sean Peisert, Dan Arnold, Nathan Hillson, Gyorgy Babnigg, Jose Manuel Marti, Christopher J. Mungall, Gregg T. Beckham, Lucas Waldburger, James Carothers, ShivShankar Sundaram, Deb Agarwal, Blake A. Simmons, Tyler Backman, Deepanwita Banerjee, Deepti Tanjore, Lavanya Ramakrishnan, and Anup Singh. Perspectives for self-driving labs in synthetic biology, 2022.
 - [42] Yuxing Fei, Bernardus Rendy, Rishi Kumar, Olympia Dartsi, Hrushikesh P. Sahasrabuddhe, Matthew J. McDermott, Zheren Wang, Nathan J. Szymanski, Lauren N. Walters, David Milsted, Yan Zeng, Anubhav Jain, and Gerbrand Ceder. Alabos: A python-based reconfigurable workflow management framework for autonomous laboratories, 2024.
 - [43] Ziqi Ni, Yahao Li, Kaijia Hu, Kunyuan Han, Ming Xu, Xingyu Chen, Fengqi Liu, Yicong Ye, and Shuxin Bai. Matpilot: an llm-enabled ai materials scientist under the framework of human-machine collaboration, 2024.
 - [44] Chemical science.
 - [45] Luis M. Antunes, Keith T. Butler, and Ricardo Grau-Crespo. Crystal structure generation with autoregressive large language modeling, 2024.
 - [46] Georgios Th. Papadopoulos, Margherita Antona, and Constantine Stephanidis. Towards open and expandable cognitive ai architectures for large-scale multi-agent human-robot collaborative learning, 2021.
 - [47] A. Gilad Kusne, Heshan Yu, Changming Wu, Huairuo Zhang, Jason Hattrick-Simpers, Brian DeCost, Suchismita Sarker, Corey Oses, Cormac Toher, Stefano Curtarolo, Albert V. Davydov, Ritesh Agarwal, Leonid A. Bendersky, Mo Li, Apurva Mehta, and Ichiro Takeuchi. On-the-fly closed-loop autonomous materials discovery via bayesian active learning, 2020.

-
- [48] Maxim Ziatdinov, Yongtao Liu, Kyle Kelley, Rama Vasudevan, and Sergei V. Kalinin. Bayesian active learning for scanning probe microscopy: from gaussian processes to hypothesis learning, 2022.
 - [49] Wenxiao Zhang, Xiangrui Kong, Conan Dewitt, Thomas Braunl, and Jin B. Hong. A study on prompt injection attack against llm-integrated mobile robotic systems, 2024.
 - [50] Sergei V. Kalinin, Maxim A. Ziatdinov, Jacob Hinkle, Stephen Jesse, Ayana Ghosh, Kyle P. Kelley, Andrew R. Lupini, Bobby G. Sumpter, and Rama K. Vasudevan. Automated and autonomous experiment in electron and scanning probe microscopy, 2021.
 - [51] Srilok Srinivasan, Rohit Batra, Duan Luo, Troy Loeffler, Sukriti Manna, Henry Chan, Liuxiang Yang, Wenge Yang, Jianguo Wen, Pierre Darancet, and Subramanian Sankaranarayanan. Machine learning the metastable phase diagram of materials, 2021.
 - [52] Siyu Liu, Tongqi Wen, Beilin Ye, Zhuoyuan Li, and David J. Srolovitz. Large language models for material property predictions: elastic constant tensor prediction and materials design, 2024.
 - [53] Utkarsh Pratiush, Hiroshi Funakubo, Rama Vasudevan, Sergei V. Kalinin, and Yongtao Liu. Scientific exploration with expert knowledge (seek) in autonomous scanning probe microscopy with active learning, 2024.
 - [54] Nicholas Roy, Ingmar Posner, Tim Barfoot, Philippe Beaudoin, Yoshua Bengio, Jeannette Bohg, Oliver Brock, Isabelle Depatie, Dieter Fox, Dan Koditschek, Tomas Lozano-Perez, Vikash Mansinghka, Christopher Pal, Blake Richards, Dorsa Sadigh, Stefan Schaal, Gaurav Sukhatme, Denis Therien, Marc Toussaint, and Michiel Van de Panne. From machine learning to robotics: Challenges and opportunities for embodied intelligence, 2021.
 - [55] Shunichi Hato and Nozomi Ogawa. Autonomous monitoring of pharmaceutical rd laboratories with 6 axis arm equipped quadruped robot and generative ai: A preliminary study, 2024.
 - [56] Felix Thelen, Lars Banko, Rico Zehl, Sabrina Baha, and Alfred Ludwig. Speeding up high-throughput characterization of materials libraries by active learning: autonomous electrical resistance measurements, 2023.
 - [57] Ganesh Sivaraman, Anand Narayanan Krishnamoorthy, Matthias Baur, Christian Holm, Marius Stan, Gabor Csányi, Chris Benmore, and Álvaro Vázquez-Mayagoitia. Machine learning inter-atomic potentials generation driven by active learning: A case study for amorphous and liquid hafnium dioxide, 2019.
 - [58] Rickard Armiento. Database-driven high-throughput calculations and machine learning models for materials design, 2019.
 - [59] Chi Chen and Shyue Ping Ong. Atomsets – a hierarchical transfer learning framework for small and large materials datasets, 2021.
 - [60] Santiago Miret, Kin Long Kelvin Lee, Carmelo Gonzales, Marcel Nassar, and Matthew Spellings. The open matsci ml toolkit: A flexible framework for machine learning in materials science, 2022.
 - [61] Martin Hangaard Hansen, José A. Garrido Torres, Paul C. Jennings, Ziyun Wang, Jacob R. Boes, Osman G. Mamun, and Thomas Bligaard. An atomistic machine learning package for surface science and catalysis, 2019.
 - [62] Justin S. Smith, Ben Nebgen, Nicholas Lubbers, Olexandr Isayev, and Adrian E. Roitberg. Less is more: sampling chemical space with active learning, 2018.
 - [63] Robin Hilgers, Daniel Wortmann, and Stefan Blügel. Application of batch learning for boosting high-throughput ab initio success rates and reducing computational effort required using data-driven processes, 2023.
 - [64] James Damewood, Jessica Karaguesian, Jaclyn R. Lunger, Aik Rui Tan, Mingrou Xie, Jiayu Peng, and Rafael Gómez-Bombarelli. Representations of materials for machine learning, 2023.

-
- [65] Kin Long Kelvin Lee, Carmelo Gonzales, Matthew Spellings, Mikhail Galkin, Santiago Miret, and Nalini Kumar. Towards foundation models for materials science: The open matsci ml toolkit, 2023.
 - [66] Haoqiang Guo, Sendong Zhao, Haochun Wang, Yanrui Du, and Bing Qin. Moltailor: Tailoring chemical molecular representation to specific tasks via text prompts, 2024.
 - [67] Leonid Kahle, Benoit Minisini, Tai Bui, Jeremy T. First, Corneliu Buda, Thomas Goldman, and Erich Wimmer. A dual-cutoff machine-learned potential for condensed organic systems obtained via uncertainty-guided active learning, 2024.
 - [68] Aanish Paruchuri, Yunfei Wang, Xiaodan Gu, and Arthi Jayaraman. Machine learning for analyzing atomic force microscopy (afm) images generated from polymer blends, 2024.
 - [69] Luis M. Antunes, Ricardo Grau-Crespo, and Keith T. Butler. Distributed representations of atoms and materials for machine learning, 2021.
 - [70] Maicon Pierre Lourenço, Hadi Zadeh-Haghghi, Jiří Hostaš, Mosayeb Naseri, Daya Gaur, Christoph Simon, and Dennis R. Salahub. Exploring quantum active learning for materials design and discovery, 2024.
 - [71] Yanjun Liu, Milena Jovanovic, Krishnanand Mallayya, Wesley J. Maddox, Andrew Gordon Wilson, Sebastian Klemenz, Leslie M. Schoop, and Eun-Ah Kim. Materials expert-artificial intelligence for materials discovery, 2023.
 - [72] Tien-Cuong Nguyen, Van-Quyen Nguyen, Van-Linh Ngo, Quang-Khoat Than, and Tien-Lam Pham. Learning hidden chemistry with deep neural networks, 2021.
 - [73] Zechen Tang, Nianlong Zou, He Li, Yuxiang Wang, Zilong Yuan, Honggeng Tao, Yang Li, Zezhou Chen, Boheng Zhao, Minghui Sun, Hong Jiang, Wenhui Duan, and Yong Xu. Improving density matrix electronic structure method by deep learning, 2024.
 - [74] Sterling G. Baird, Marianne Liu, Hasan M. Sayeed, and Taylor D. Sparks. Data-driven materials discovery and synthesis using machine learning methods, 2022.
 - [75] Bo Hu, Siyu Liu, Beilin Ye, Yun Hao, and Tongqi Wen. A multi-agent framework for materials laws discovery, 2024.
 - [76] Armin Sahinovic and Benjamin Geisler. Active learning and element embedding approach in neural networks for infinite-layer versus perovskite oxides, 2021.
 - [77] Rajendra P. Joshi, Jesse Eickholt, Liling Li, Marco Fornari, Veronica Barone, and Juan E. Peralta. Machine learning the voltage of electrode materials in metal-ion batteries, 2019.
 - [78] Victor Trinquet, Matthew L. Evans, Cameron J. Hargreaves, Pierre-Paul De Breuck, and Gian-Marco Rignanese. Optical materials discovery and design with federated databases and machine learning, 2024.
 - [79] Jihua Chen, Yue Yuan, Amir Koushyar Ziabari, Xuan Xu, Honghai Zhang, Panagiotis Christakopoulos, Peter V. Bonnesen, Ilia N. Ivanov, Panchapakesan Ganesh, Chen Wang, Karen Patino Jaimes, Guang Yang, Rajeev Kumar, Bobby G. Sumpter, and Rigoberto Advincula. Ai for manufacturing and healthcare: a chemistry and engineering perspective, 2024.
 - [80] Rongzhi Dong, Nihang Fu, dirisuriya M. D. Siriwardane, and Jianjun Hu. Generative design of inorganic compounds using deep diffusion language models, 2023.
 - [81] Isaac Filella-Merce, Alexis Molina, Marek Orzechowski, Lucía Díaz, Yang Ming Zhu, Julia Vilalta Mor, Laura Malo, Ajay S Yekkirala, Soumya Ray, and Victor Guallar. Optimizing drug design by merging generative ai with active learning frameworks, 2023.
 - [82] Chi Chen, Dan Thien Nguyen, Shannon J. Lee, Nathan A. Baker, Ajay S. Karakoti, Linda Lauw, Craig Owen, Karl T. Mueller, Brian A. Bilodeau, Vijayakumar Murugesan, and Matthias Troyer. Accelerating computational materials discovery with artificial intelligence and cloud high-performance computing: from large-scale screening to experimental validation, 2024.

-
- [83] Saleem Abdul Fattah Ahmed Al Dajani and David Keyes. Constructing artificial life and materials scientists with accelerated ai using deep andersonn, 2024.
 - [84] Sebastien Röcken and Julija Zavadlav. Accurate machine learning force fields via experimental and simulation data fusion, 2023.
 - [85] K. T. Schütt, M. Gastegger, A. Tkatchenko, K. R. Müller, and R. J. Maurer. Unifying machine learning and quantum chemistry – a deep neural network for molecular wavefunctions, 2019.
 - [86] Julia Westermayr, Michael Gastegger, Kristof T. Schütt, and Reinhard J. Maurer. Perspective on integrating machine learning into computational chemistry and materials science, 2021.
 - [87] Anton V. Sinit斯基 and Vijay S. Pande. Physical machine learning outperforms "human learning" in quantum chemistry, 2020.
 - [88] Lenz Fiedler, Karan Shah, Michael Bussmann, and Attila Cangi. A deep dive into machine learning density functional theory for materials science and chemistry, 2022.
 - [89] Marco Fronzi, Mutaz Abu Ghazaleh, Olexandr Isayev, David A. Winkler, Joe Shapter, and Michael J. Ford. Impressive computational acceleration by using machine learning for 2-dimensional super-lubricant materials discovery, 2020.
 - [90] Bing Huang, Guido Falk von Rudorff, and O. Anatole von Lilienfeld. Towards self-driving laboratories: The central role of density functional theory in the ai age, 2023.
 - [91] Microsoft Research AI4Science and Microsoft Azure Quantum. The impact of large language models on scientific discovery: a preliminary study using gpt-4, 2023.
 - [92] Konstantin Gubaev, Evgeny V. Podryabinkin, Gus L. W. Hart, and Alexander V. Shapeev. Accelerating high-throughput searches for new alloys with active learning of interatomic potentials, 2018.
 - [93] Riley J. Hickman, Florian Häse, Loïc M. Roch, and Alán Aspuru-Guzik. Gemini: Dynamic bias correction for autonomous experimentation and molecular simulation, 2021.
 - [94] Maxim Ziatdinov, Yongtao Liu, Anna N. Morozovska, Eugene A. Eliseev, Xiaohang Zhang, Ichiro Takeuchi, and Sergei V. Kalinin. Hypothesis learning in automated experiment: Application to combinatorial materials libraries, 2022.
 - [95] Troy D. Loeffler, Tarak K. Patra, Henry Chan, Mathew Cherukara, and Subramanian K. R. S. Sankaranarayanan. Active learning the coarse-grained energy landscape for water clusters from sparse training data, 2019.
 - [96] Sokseihha Muy, Conrad Johnston, and Nicola Marzari. Aiida-defects: An automated and fully reproducible workflow for the complete characterization of defect chemistry in functional materials, 2023.
 - [97] Adarsh Dave, Jared Mitchell, Sven Burke, Hongyi Lin, Jay Whitacre, and Venkatasubramanian Viswanathan. Autonomous optimization of nonaqueous battery electrolytes via robotic experimentation and machine learning, 2021.
 - [98] Abbi Abdel-Rehim, Hector Zenil, Oghenejokpeme Orhobor, Marie Fisher, Ross J. Collins, Elizabeth Bourne, Gareth W. Fearnley, Emma Tate, Holly X. Smith, Larisa N. Soldatova, and Ross D. King. Scientific hypothesis generation by a large language model: Laboratory validation in breast cancer treatment, 2024.
 - [99] Guy Hacohen and Daphna Weinshall. How to select which active learning strategy is best suited for your specific problem and budget, 2023.
 - [100] Mohnish Harwani, Juan C. Verduzco, Brian H. Lee, and Alejandro Strachan. Accelerating active learning materials discovery with fair data and workflows: a case study for alloy melting temperatures, 2024.

-
- [101] Mario Boley, Felix Luong, Simon Teshuva, Daniel F Schmidt, Lucas Foppa, and Matthias Scheffler. From prediction to action: Critical role of performance estimation for machine-learning-driven materials discovery, 2023.
 - [102] Xiangyun Lei, Edward Kim, Viktoriia Baibakova, and Shijing Sun. Lessons in reproducibility: Insights from nlp studies in materials science, 2023.
 - [103] Bhavya Kailkhura, Brian Gallagher, Sookyoung Kim, Anna Hiszpanski, and T. Yong-Jin Han. Reliable and explainable machine learning methods for accelerated material discovery, 2019.
 - [104] Felipe Oviedo, Juan Lavista Ferres, Tonio Buonassisi, and Keith Butler. Interpretable and explainable machine learning for materials science and chemistry, 2021.
 - [105] Shufeng Kong, Dan Guevarra, Carla P. Gomes, and John M. Gregoire. Materials representation and transfer learning for multi-property prediction, 2021.
 - [106] Sheng Gong, Shuo Wang, Taishan Zhu, Yang Shao-Horn, and Jeffrey C. Grossman. Multi-modal machine learning for materials science: composition-structure bimodal learning for experimentally measured properties, 2023.
 - [107] Tianhao Chen, Pengbo Xu, and Pengbo Xu. Botfip-llm: An enhanced multimodal scientific computing framework leveraging knowledge distillation from large language models, 2024.
 - [108] Mathias S. Jørgensen, Henrik L. Mortensen, Søren A. Meldgaard, Esben L. Kolsbjerg, Thomas L. Jacobsen, Knud H. Sørensen, and Bjørk Hammer. Atomistic structure learning, 2019.
 - [109] Lingyu Kong, Jielan Li, Lixin Sun, Han Yang, Hongxia Hao, Chi Chen, Nongnuch Artrith, Jose Antonio Garrido Torres, Ziheng Lu, and Yichi Zhou. Overcoming the size limit of first principles molecular dynamics simulations with an in-distribution substructure embedding active learner, 2023.
 - [110] James Dean, Matthias Scheffler, Thomas A. R. Purcell, Sergey V. Barabash, Rahul Bhowmik, and Timur Bazhirov. Interpretable machine learning for materials design, 2021.
 - [111] Viggo Moro, Charlotte Loh, Rumen Dangovski, Ali Ghorashi, Andrew Ma, Zhuo Chen, Samuel Kim, Peter Y. Lu, Thomas Christensen, and Marin Soljačić. Multimodal learning for materials, 2024.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

www.SurveyX.Cn