

Doi:10.32604/cmc.2025.067427

REVIEW





Large Language Model-Driven Knowledge Discovery for Designing Advanced Micro/Nano Electrocatalyst Materials

Ying Shen¹, Shichao Zhao¹, Yanfei Lv¹, Fei Chen¹, Li Fu^{1,*} and Hassan Karimi-Maleh^{2,*}

¹College of Materials and Environmental Engineering, Hangzhou Dianzi University, Hangzhou, 310018, China
²The Quzhou Affiliated Hospital of Wenzhou Medical University, Quzhou People's Hospital, Quzhou, 324000, China
*Corresponding Authors: Li Fu. Email: fuli@hdu.edu.cn; Hassan Karimi-Maleh. Email: hassan@wmu.edu.cn
Received: 03 May 2025; Accepted: 12 June 2025; Published: 03 July 2025

ABSTRACT: This review presents a comprehensive and forward-looking analysis of how Large Language Models (LLMs) are transforming knowledge discovery in the rational design of advanced micro/nano electrocatalyst materials. Electrocatalysis is central to sustainable energy and environmental technologies, but traditional catalyst discovery is often hindered by high complexity, fragmented knowledge, and inefficiencies. LLMs, particularly those based on Transformer architectures, offer unprecedented capabilities in extracting, synthesizing, and generating scientific knowledge from vast unstructured textual corpora. This work provides the first structured synthesis of how LLMs have been leveraged across various electrocatalysis tasks, including automated information extraction from literature, text-based property prediction, hypothesis generation, synthesis planning, and knowledge graph construction. We comparatively analyze leading LLMs and domain-specific frameworks (e.g., CatBERTa, CataLM, CatGPT) in terms of methodology, application scope, performance metrics, and limitations. Through curated case studies across key electrocatalytic reactions-HER, OER, ORR, and CO₂RR-we highlight emerging trends such as the growing use of embedding-based prediction, retrieval-augmented generation, and fine-tuned scientific LLMs. The review also identifies persistent challenges, including data heterogeneity, hallucination risks, lack of standard benchmarks, and limited multimodal integration. Importantly, we articulate future research directions, such as the development of multimodal and physics-informed MatSci-LLMs, enhanced interpretability tools, and the integration of LLMs with selfdriving laboratories for autonomous discovery. By consolidating fragmented advances and outlining a unified research roadmap, this review provides valuable guidance for both materials scientists and AI practitioners seeking to accelerate catalyst innovation through large language model technologies.

KEYWORDS: Large language models; electrocatalysis; nanomaterials; knowledge discovery; materials design; artificial intelligence; natural language processing

1 Introduction

Advanced electrocatalyst materials, particularly those engineered at the micro- and nano-scale, stand at the forefront of solutions addressing critical global challenges in sustainable energy and environmental stewardship [1]. Their importance is underscored by their central role in enabling key technologies such as fuel cells for clean power generation [2], water electrolyzers for green hydrogen production [3], systems for converting greenhouse gases like CO_2 into valuable chemicals via CO_2 reduction [4], and biomass/waste valorization through electrochemical upgrading of compounds such as waste glycerol [5], bio-oil [6], and waste cooking oil [7] into value-added fuels and chemicals [8,9]. The efficiency, selectivity, and durability of these electrochemical processes are fundamentally dictated by the performance of the



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

electrocatalyst employed. Consequently, the development of catalysts exhibiting superior activity, enhanced selectivity towards desired products, and robust stability under demanding operational conditions remains a paramount objective in materials science and chemical engineering research. Micro/nano structuring improves catalytic performance by increasing surface area and exposing more active sites, such as edges and defects. Despite decades of research, the traditional path to discovering and optimizing electrocatalysts is often arduous and inefficient [10–12]. Traditional discovery often depends on chemical intuition, time-consuming experiments, and DFT-based simulations [13–15]. These approaches face significant hurdles. The combinatorial chemical space encompassing potential catalyst compositions, structures, and morphologies is astronomically vast, rendering exhaustive exploration practically impossible [12,14,16]. Furthermore, the relationship between a material's atomic-level structure, its electronic properties, and its ultimate catalytic performance is exceptionally complex and often non-intuitive, making rational design challenging [11,12,17]. Although fundamental studies using model systems are insightful, they often face the "materials gap" and "pressure gap". This means findings from idealized conditions, such as single crystals under ultrahigh vacuum, may not apply to real-world nanoparticle catalysts. The complex structure of nanocatalysts often exceeds the ability of current tools to identify active sites or explain reaction mechanisms.

Recognizing these limitations, the scientific community is increasingly embracing a data-driven paradigm, often referred to as the "fourth paradigm of science", to accelerate materials discovery [13,18–20]. This approach leverages the power of Artificial Intelligence (AI) and Machine Learning (ML) techniques to analyze large datasets, uncover hidden patterns, predict material properties, and guide experimental efforts [10,17,20–23]. The exponential growth in both computational power and the availability of materials data, generated through high-throughput simulations and automated experiments, has fueled this transition. ML models have demonstrated success in predicting catalyst properties and screening candidate materials based on structured datasets derived from simulations or experiments [17,20,23]. These traditional ML approaches typically rely on numerical descriptors or graph-based representations derived from well-structured input data, such as atomic coordinates or engineered features. While powerful, their utility is often limited by the availability of high-quality, labeled datasets and their inability to fully leverage the rich but unstructured scientific knowledge embedded in textual sources.

In contrast, LLMs represent a fundamentally different class of AI systems. Built on Transformer architectures and trained on vast corpora of unstructured text, LLMs are capable of understanding and generating human language, making them especially suited for processing scientific literature. Unlike conventional ML models that operate primarily on structured numerical input, LLMs can perform tasks such as text mining, summarization, question answering, and entity-relation extraction directly from natural language. These abilities enable LLMs to unlock and synthesize latent scientific insights from previously untapped textual resources, offering a powerful complement to existing data-driven catalyst design strategies [24]. Transformer architecture and trained on massive corpora of scientific literature, LLMs excel at tasks such as text mining, summarization, question answering, and entity-relation extraction. These capabilities are particularly relevant to catalysis research, which is characterized by extensive unstructured knowledge dispersed across publications, patents, and technical reports [25]. Recent work has demonstrated the potential of LLMs to extract synthesis protocols and performance metrics from catalyst literature [26], build structured databases from text, and even assist in predicting catalyst properties based on natural language descriptions. For instance, LLM-based models such as SciBERT and CatBERTa have been used to extract materials and adsorption energy data with high accuracy, and domain-specific models like CataLM have been fine-tuned specifically for electrocatalytic materials. Furthermore, generative models such as CatGPT have been trained to propose novel catalyst compositions and reaction mechanisms, enabling hypothesis generation at scale. These developments illustrate how LLMs are being increasingly integrated into catalyst

design pipelines—either as standalone tools for literature analysis or as components in multi-agent or closed-loop discovery systems. By processing the language of science itself, LLMs are beginning to serve not only as data extraction engines, but as active collaborators in the generation of knowledge and the acceleration of catalyst innovation. This is vital because most scientific knowledge exists in unstructured text, not in structured databases [19,27–29]. Traditional ML methods typically struggle to directly utilize this vast textual resource, often requiring laborious manual data extraction or sophisticated, domain-specific Natural Language Processing (NLP) pipelines. LLMs, however, offer a more direct route to harnessing this knowledge, acting potentially as "second brains" for researchers by processing and synthesizing information from literature at an unprecedented scale [29].

This review focuses specifically on the application of LLMs to drive knowledge discovery for the purpose of designing advanced micro/nano electrocatalyst materials. Fig. 1 provides a schematic roadmap outlining the core roles LLMs play in this context. LLMs assist in several key tasks: extracting synthesis protocols and performance data from the literature, predicting catalytic properties using semantic understanding, generating hypothetical structures or mechanisms, and integrating knowledge via summarization or knowledge graphs. By capturing the entire knowledge discovery cycle—ranging from data extraction to hypothesis formulation—the figure visually reinforces the manuscript's central thesis that LLMs are emerging as comprehensive enablers in catalyst design.



Figure 1: Schematic overview of how LLMs can be applied for knowledge discovery in electrocatalyst design

2 Foundational Concepts: Advanced Electrocatalysis and LLMs

A comprehensive understanding of the LLM-driven approach to electrocatalyst design necessitates familiarity with the fundamental aspects of both the target materials and the AI tools being employed. This section provides foundational concepts for advanced micro/nano electrocatalyst materials and LLMs, setting the stage for subsequent discussions on their intersection.

2.1 Advanced Micro/Nano Electrocatalyst Materials

Electrocatalysis is fundamentally the process by which the rate of an electrochemical reaction occurring at an electrode-electrolyte interface is enhanced through the action of a catalytic material. These reactions are central to numerous energy conversion and storage technologies, including fuel cells, electrolyzers, and metal-air batteries, as well as environmental applications like CO_2 conversion and pollutant degradation [30]. Advanced electrocatalysts are engineered with unique structural or compositional features, often at the micro/nanoscale (1–100 nm). Nanostructuring offers multiple benefits. It increases the surface area available to reactants, introduces quantum confinement and modified electronic properties, and exposes low-coordination sites like edges and corners that enhance catalytic activity [11,31]. These benefits lead to higher activity, better selectivity, and improved atom efficiency, making nano-electrocatalysts valuable for energy and environmental applications.

The field encompasses a diverse array of material types, many of which are being explored using LLM-driven approaches. Recent years have seen substantial advances in the rational design of micro/nano electrocatalysts that exhibit improved intrinsic activity, selectivity, and stability. For example, single-atom catalysts (SACs) with isolated metal atoms anchored on nitrogen-doped carbon matrices have demonstrated record-low overpotentials for HER and OER [11,32]. Transition metal dichalcogenides (TMDs), especially MoS₂ and WS₂, have been modified through phase engineering and defect creation to enhance ORR and CO₂RR performance [15,30,33]. High-entropy alloys (HEAs) have emerged as tunable electrocatalyst platforms with compositional flexibility, showing promising results for multi-functional catalysis [12,17,34,35]. Metal-organic frameworks (MOFs) and their derivatives are now widely investigated for their hierarchical porosity and tunable active sites, often serving as precursors for metal-N-C catalysts. Porous and hollow nanostructures, such as yolk-shell and nanoframe morphologies, are increasingly favored for their ability to enhance mass transport and facilitate intermediate desorption. Table 1 synthesizes recent advances across various categories of micro/nano electrocatalyst materials, providing a comparative overview of their unique structural features, target reactions, representative materials, and performance benchmarks. Metal nanoparticles, often based on noble metals like Pt, Pd, or Au, or non-noble transition metals, can be used either unsupported or dispersed on conductive supports like carbon materials, oxides, or nitrides [11,31]. Their catalytic properties are highly sensitive to size, shape, and composition. One-dimensional (1D) nanostructures, such as nanowires, nanorods, and nanotubes, offer directed electron transport pathways and high surface areas [15]. Two-dimensional (2D) materials, including graphene and its analogues, TMDs, MXenes, and the more recently explored metallenes (ultrathin metal nanosheets), provide extremely high surface exposure and unique electronic properties stemming from their reduced dimensionality [15,30,33]. High-entropy materials (HEMs), including HEAs and oxides, contain five or more main elements in nearly equal ratios. Their disordered atomic structures produce unique synergistic effects and a vast compositional space for catalyst design [12,17,34,35]. SACs represent the ultimate limit of atom efficiency, where individual metal atoms are dispersed on a support material, providing well-defined active sites and potentially unique catalytic behavior distinct from nanoparticles [11,32]. Other important classes include nanostructured metal oxides, sulfides, phosphides, and (oxy)hydroxides, which are often investigated for reactions like oxygen evolution [11,36], as well as porous materials like MOFs that can serve as catalyst precursors or platforms [30,37].

Catalyst type	Key features	Target reaction (s)	Representative materials	Performance highlights	Refs.
Single-Atom	Atomically	HER, OER, ORR,	Fe–N–C, Co–N–C	Overpotential <	[11,38]
Catalysts (SACs)	dispersed metals on supports	CO ₂ RR		50 mV (HER); High TOF	
Transition Metal	Layered 2D	HER, ORR,	MoS_2 , WS_2	HER onset potential	[15,36,39]
Dichalcogenides (TMDs)	structures, defect engineering	CO ₂ RR		~70 mV	
High-Entropy	Multicomponent	HER, ORR	PtPdNiFeCo,	Adjustable binding	[12,17,40,41]
Alloys (HEAs)	equiatomic alloys		AgPdRu	energies, low η_{10}^*	

Table 1: Summary of recent advances in micro/nano electrocatalyst materials

(Continued)

Table 1 (continued)

Catalyst type	Key features	Target reaction (s)	Representative materials	Performance highlights	Refs.
MOFs/Derived carbon composites	Hierarchical porosity, tunable active sites	CO ₂ RR, OER	ZIF-8 derived Co-N-C	FE(CO) > 90% (CO ₂ RR)	[36,42]
Nanoframes & hollow structures	Enhanced surface exposure, mass transport	OER, ORR	PtNi nanoframes, Co ₃ O4 hollow spheres	High stability, η ₁₀ < 300 mV (OER)	[11,43]

Note: η_{10}^* refers to the overpotential required to achieve a current density of 10 mA·cm⁻², which is a standard metric for assessing electrocatalytic activity. Lower η_{10} values indicate higher catalytic efficiency, as less energy is required to drive the target electrochemical reaction at a practical current density.

Evaluating the effectiveness of these diverse electrocatalysts requires standardized key performance indicators (KPIs). Catalyst activity is often measured by overpotential (η), the extra voltage needed to reach 10 mA cm⁻² current density. Lower overpotential signifies higher activity. Other key metrics include onset potential, current density, Tafel slope for mechanism insight, and TOF per active site, though defining active sites can be difficult [16,38]. Selectivity, crucial for reactions producing multiple products like CO₂RR, is measured by the Faradaic efficiency (FE), representing the percentage of electrons contributing to the formation of the desired product [39,40]. Stability reflects a catalyst's ability to retain performance and structure during long-term use, often tracked by potential or current changes [38]. These KPIs are essential for comparing different materials and guiding the design process.

These advanced electrocatalysts are indispensable for driving key electrochemical reactions central to sustainable technologies. Prominent examples include the hydrogen evolution reaction (HER) and oxygen evolution reaction (OER) involved in water splitting for hydrogen production; the oxygen reduction reaction (ORR) and hydrogen oxidation reaction (HOR) critical for fuel cells; the carbon dioxide reduction reaction (CO₂RR) for converting CO₂ into fuels and chemicals; the Nitrogen Reduction Reaction (NRR) for ammonia synthesis; and the oxidation of fuels like ethanol or methanol in direct fuel cells [17,41,43]. However, the rational design of these high-performance micro/nano electrocatalysts remains a formidable task due to several intrinsic challenges [42]. The sheer complexity of these systems, involving the intricate interplay between composition, size, shape, surface structure, defects, and support interactions, makes predicting behavior difficult [11,12]. Establishing clear structure-property relationships that link these nanoscale features to macroscopic catalytic performance (activity, selectivity, stability) is often elusive [12,17]. The Sabatier principle states that effective catalysis requires moderate binding: too weak fails to activate, too strong poisons the surface [1]. Finding materials that strike this delicate balance for complex multi-step reactions is challenging. Precise synthesis control to reproducibly fabricate nanomaterials with desired structural attributes remains a significant hurdle. Characterizing these materials under real conditions is difficult, limiting understanding of their behavior and true active sites [38]. These challenges collectively highlight the need for advanced tools, such as LLMs, capable of navigating complexity, extracting knowledge from vast datasets, and accelerating the design cycle.

2.2 LLMs for Scientific Discovery

LLMs represent a significant advancement in artificial intelligence, characterized by their massive scale (billions of parameters) and their ability to process and generate human-like text [24]. Their development has been fueled by breakthroughs in neural network architectures, particularly the Transformer, increased computational power, and the availability of vast amounts of training data.

The core principles underlying most modern LLMs involve several key components. The transformer architecture is fundamental, replacing the sequential processing of older recurrent neural networks (RNNs) with parallel processing enabled by attention mechanisms. Self-attention allows the model to weigh the importance of different words or tokens within an input sequence relative to each other, capturing contextual relationships and long-range dependencies effectively [24]. Variations like sparse attention, flash attention, and multi-query attention have been developed to improve efficiency and handle longer sequences. LLMs typically undergo a two-stage training process. First is pre-training, where the model learns general language understanding and world knowledge by training on enormous, diverse text corpora (often terabytes of data) using self-supervised objectives like predicting the next word in a sequence or filling in masked words. This phase imbues the model with broad linguistic competence. Second is fine-tuning, where the pre-trained model is further trained on smaller, more specific datasets tailored to particular tasks or domains. This adaptation stage includes instruction-tuning (training on instruction-response pairs to follow commands) and alignment-tuning (often using reinforcement learning from human feedback, RLHF) to make the model more helpful, honest, and harmless [44]. A remarkable aspect of LLMs is their emergent abilities-capabilities like complex reasoning, planning, few-shot learning (performing tasks with only a few examples), and zero-shot learning (performing tasks without examples) that appear as model scale increases and were not explicitly programmed [45].

These underlying principles grant LLMs a suite of capabilities highly relevant to scientific knowledge discovery, particularly from the vast body of scientific literature. IE is a key capability, allowing LLMs to identify and extract specific pieces of information from unstructured text, such as experimental parameters, material compositions, property values, and synthesis steps [19,24,29,46,47]. This includes tasks like Named Entity Recognition (NER) to identify relevant terms and Relation Extraction to understand how they are connected [19]. Knowledge synthesis and summarization enable LLMs to condense information from lengthy articles or multiple sources, generate literature reviews, and identify key findings, helping researchers grapple with information overload. Question answering (Q&A) allows users to pose specific questions and receive answers based on the LLM's internal knowledge or provided documents. Techniques like retrievalaugmented generation (RAG), where the LLM retrieves relevant information from an external database or corpus before generating an answer, enhance factual grounding and accuracy [25,48]. LLMs can also contribute to hypothesis generation by analyzing existing literature to identify knowledge gaps, inconsistencies, or potential correlations that might suggest new research avenues or material candidates [49,50]. Furthermore, many LLMs possess strong code generation abilities, capable of writing code snippets for data analysis, simulation setup, or even controlling automated laboratory equipment based on natural language prompts [51,52]. Finally, their core text generation capability can assist researchers in drafting manuscripts, reports, or documentation [47,53]. Fig. 2 serves as a conceptual framework illustrating the multilayered architecture of LLMs-from transformer-based attention mechanisms to fine-tuning strategies-and maps these capabilities to practical functions in scientific discovery. The figure highlights key tasks such as named entity recognition, summarization, question answering, and code generation, all of which are crucial for mining and synthesizing information from scientific literature. This visual contextualizes how LLMs act as both knowledge extractors and generators, thereby facilitating end-to-end integration into research workflows, including literature analysis, database construction, and experimental planning.



Figure 2: Schematic illustration of how LLMs, through their core architecture and training processes, enable key capabilities such as IE, summarization, and code generation, thereby supporting various stages of scientific knowledge discovery, from literature analysis to hypothesis formulation and manuscript preparation

Collectively, these capabilities position LLMs as powerful tools for a new knowledge discovery paradigm in materials science and other scientific fields. Traditional knowledge discovery often relies on structured databases, which cover only a fraction of published knowledge, or manual literature surveys, which are slow and limited in scope. LLMs offer the potential to bridge this gap by directly processing and interpreting the primary source of scientific knowledge—the unstructured text of publications—at scale. By extracting, synthesizing, and reasoning over this vast textual information, LLMs can potentially accelerate the identification of structure-property relationships, suggest novel materials, optimize experimental procedures, and ultimately speed up the cycle of scientific discovery in complex fields like electrocatalyst design.

3 LLM Methodologies in Micro/Nano Electrocatalyst Research

The application of LLMs in the domain of micro/nano electrocatalyst research is rapidly evolving, with various methodologies being developed and tested to leverage their unique capabilities. These methodologies broadly fall into categories focused on extracting existing knowledge, predicting properties based on that knowledge, generating new hypotheses or plans, and synthesizing information across the literature.

3.1 IE from Electrocatalysis Literature

In the pursuit of accelerating electrocatalyst discovery, one of the most formidable challenges lies in the accessibility of materials data. Unlike structured databases, a significant proportion of valuable knowledge remains embedded in the unstructured text of scientific publications, patents, and supplementary materials. IE using LLMs has emerged as a transformative solution to this problem. By converting free-text content into structured, machine-readable formats, LLMs enable the creation of comprehensive datasets that capture key experimental and material parameters—thereby forming the foundation for downstream tasks such as property prediction, hypothesis generation, and synthesis planning [54,55]. The overarching goal is to create comprehensive, machine-readable datasets detailing materials, synthesis methods, experimental conditions, properties, and performance metrics from the vast electrocatalysis literature [54].

NER and relation extraction are fundamental tasks in this process. NER involves identifying specific entities within the text, such as catalyst materials (e.g., 'Pt nanoparticles', 'Cu-based alloys', 'perovskite oxides'), precursors, reaction products (e.g., 'H₂', 'CO', 'CH₄'), performance metrics (e.g., FE, overpotential, current density, Tafel slope), experimental parameters (e.g., temperature, pressure, electrolyte composition, potential range), synthesis operations (e.g., 'calcination', 'electrodeposition'), and characterization techniques [19,40,54,55]. Relation Extraction then identifies the semantic links between these entities, for example, connecting a specific catalyst material to its measured FE for a particular product under defined conditions.

The effectiveness of IE in electrocatalysis relies heavily on the capabilities of the underlying LLMs. Early efforts employed general-purpose models such as GPT-01 and GPT-40, which—when guided by carefully designed prompts—proved capable of extracting synthesis parameters and performance metrics with high accuracy from large corpora [37,46,54]. Examples include SciBERT [40], MatSciBERT [44,54], and CataLM [44], which is specifically tailored for electrocatalytic materials using the Vicuna-13B model as a base and trained on domain literature and expert annotations [44]. Comparative studies have shown promising results; for instance, GPT-4 outperformed a rule-based method (ChemDataExtractor) in extracting band gap information from materials science literature, demonstrating better handling of complex material names and interdependency resolution, although weaknesses in hallucination and identifying specific value types were noted [46]. In the context of CO₂ reduction electrocatalysis, a framework combining BERT embeddings with a BiLSTM-CRF architecture showed strong performance in recognizing key entities [56].

To enhance accuracy and mitigate issues like hallucination, various techniques are employed. Prompt engineering involves carefully crafting the input query to elicit the desired structured output from the LLM [37,46,54]. Fine-tuning adapts the model's parameters using domain-specific labeled data [44]. RAG approaches, often utilizing vector databases built from literature embeddings (e.g., using SciBERT), retrieve relevant context before generation, grounding the LLM's output and improving factual accuracy [25,54]. The ChatExtract method employs a conversational approach, using follow-up prompts to verify extracted data and introduce uncertainty checks, achieving high precision and recall for materials data extraction with models like GPT-4 [36].

These extraction efforts are enabling the creation of valuable resources. Several projects have focused on building specialized corpora, such as the benchmark and extended corpora for the CO₂ electrocatalytic reduction process, containing thousands of manually verified or automatically extracted records detailing materials, methods, products, efficiencies, and conditions [54]. Such corpora serve as vital training and evaluation datasets for future NLP models in the field.

Beyond simple entity and property extraction, LLMs are being used to parse complex experimental synthesis procedures. This involves identifying the sequence of actions (e.g., mixing, heating, washing, annealing), associated parameters (temperature, time, concentration, atmosphere), starting materials, and target products described in narrative text [54,57,58]. Sequence-to-sequence Transformer models have been developed to convert these unstructured descriptions into structured "action sequences", providing a machine-readable representation of the synthesis protocol that can be used for analysis, comparison, or even guiding automated synthesis platforms [54]. Examples include extracting synthesis parameters for inorganic materials [59] and MOFs [37].

While most current efforts focus on text, the challenge of extracting information from multimodal sources like tables and figures is recognized [27]. Early work includes extracting data from tables in conjunction with text, with specialized models like MaTableGPT emerging [58]. The future likely involves multimodal LLMs capable of interpreting images (plots, micrographs), tables, and text simultaneously [25,42]. LLM-driven IE is not simply a technical exercise in data extraction; it is a foundational

enabler for the entire AI-driven catalyst discovery pipeline. From generating structured corpora and parsing synthesis protocols to integrating text with knowledge graphs and multimodal inputs, IE continues to evolve into a pivotal capability that underpins knowledge accessibility in electrocatalysis research. The narrative arc of these efforts demonstrates a clear trajectory: from isolated data extraction to the construction of intelligent, interconnected research ecosystems.

3.2 Property Prediction and Structure-Property Relationships

A central goal in materials science is to predict the properties of a material based on its composition and structure. LLMs are introducing novel approaches to this challenge in electrocatalysis, moving beyond traditional methods that rely solely on explicit structural inputs or computationally expensive simulations [60]. These new methodologies leverage the ability of LLMs to understand and process textual information or to generate meaningful numerical representations (embeddings) from scientific language.

One promising direction involves text-based property prediction, where models predict properties directly from natural language descriptions of the material or catalytic system. CatBERTa, a model based on the RoBERTa Transformer encoder, exemplifies this approach [61]. CatBERTa, a domain-specific adaptation of the RoBERTa Transformer encoder, exemplifies the use of text-based property prediction in electrocatalysis. By encoding structured textual representations of adsorption systems (e.g., catalyst surface facets, adsorbates, site types), it enables semantic learning without needing explicit atomic coordinates. The model demonstrated comparable or superior performance to conventional GNNs like CGCNN on benchmark datasets such as OC20. A particularly impactful finding was CatBERTa's ability to minimize systematic error when evaluating energy differences among chemically similar systems, making it useful for highthroughput screening. However, its reliance on well-formatted text input can limit applicability in cases where such structured language representations are unavailable. It takes human-interpretable text strings as input, which can include information like the chemical symbols of the adsorbate and bulk catalyst, the crystal facet's Miller index, the type of adsorption site, and potential atomic properties. CatBERTa is fine-tuned to predict adsorption energies, a key descriptor for catalytic activity [61]. Fig. 3 illustrates the architecture and operational flow of CatBERTa, a domain-specific LLM fine-tuned for predicting adsorption energies in catalysis. The conversion of structural descriptors into textual inputs enables the model to derive embeddings using a Transformer encoder, which is then processed via a regression head to output predicted energy values. This framework is particularly powerful because it leverages semantic representations of material features, enabling effective prediction even in cases where atomic coordinates may be incomplete or unavailable. The visual aids in understanding the interpretability advantage and prediction mechanics of text-based LLM models in materials science. Its performance has been shown to be comparable to some established Graph Neural Networks (GNNs) like CGCNN and SchNet, particularly on certain subsets of data, achieving Mean Absolute Errors (MAE) in the range of 0.35-0.82 eV depending on the dataset and input features. A notable strength of CatBERTa is its ability to significantly cancel systematic errors when predicting energy differences between chemically similar systems, outperforming GNNs in this aspect. This suggests its utility in comparative studies and screening. Other work has also explored using text descriptions generated by tools like Robocrystallographer as input for transformer models (BERT, MatBERT) to classify materials based on properties like formation energy and band gap, achieving high accuracy [60].



Figure 3: Overview of CatBERTa: (a) Transformation of structural data into a textual format; The structural data undergoes conversion into two types of textual inputs: strings and descriptions; (b) Visualization of the fine-tuning process. The embedding from the special token "<s>" is input to the regression head, comprising a linear layer and an activation layer; (c) Illustration of the Transformer encoder and a multihead attention mechanism [61]

Another approach utilizes embedding-based prediction. This involves representing materials, properties, or concepts as dense numerical vectors (embeddings) learned by NLP models trained on large text corpora. The relationships between these vectors can then be exploited for prediction. For instance, one study trained a Word2Vec model on materials science abstracts and used the cosine similarity between material composition vectors and property term vectors (e.g., 'conductivity', 'dielectric') as objectives for Pareto optimization [14]. This allowed the prediction and screening of candidate electrocatalyst compositions (e.g., Ag-Pd-Pt, Ag-Pd-Ru) for specific reactions like HER, OER, and ORR, purely based on latent knowledge extracted from text, with predictions matching experimental activity trends well [14]. Similarly, another approach combined word embeddings derived from literature (capturing semantic information) with graph embeddings derived from a constructed knowledge graph (capturing structural relationships) [40]. This hybrid embedding was fed into a deep learning model to successfully predict the FE of Cu-based catalysts for CO₂ reduction, demonstrating the power of integrating textual semantics with structured knowledge (Fig. 4) [40].



Figure 4: (a) (Top) Stacked histograms of various Cu-based electrocatalysts in articles published in the last dozen years. (Bottom) Stacked histograms of the percentage of Cu-based electrocatalysts in articles normalized by year; (b) Overall representation of materials, products, and methods for CO_2 reduction. The size of the balls indicates the number of corresponding papers; (c) Alluvial plot showing the development of Cu-based alloy electrocatalysts across the last 30 years [40]

A potential advantage of text-based LLM approaches is interpretability. Unlike many complex ML models often termed "black boxes", text-based models offer avenues for understanding the prediction process. The attention mechanisms inherent in Transformer models like CatBERTa can be analyzed to reveal which parts of the input text (specific words or tokens, such as those related to the adsorbate or interacting atoms) the model focuses on when making a prediction [61]. This provides insights into the features the model deems important. Furthermore, using human-readable text descriptions as the primary input representation itself enhances interpretability, allowing researchers to more easily relate the model's input to their domain knowledge [60]. This contrasts with GNNs where interpretation often relies on analyzing learned graph features or using post-hoc methods like SHAP or LIME [17]. When compared with GNNs and traditional ML, LLM-based property prediction offers distinct characteristics. GNNs excel at capturing fine-grained structural information when precise atomic coordinates are available [61,62]. Traditional descriptor-based ML relies on carefully engineered features derived from physical or chemical principles [22,23]. LLMs, particularly text-based ones, operate on a different modality, leveraging semantic understanding and contextual information from language. This allows them to potentially incorporate qualitative knowledge, handle incomplete structural information, and utilize human-interpretable inputs. However, they might lack the explicit structural resolution of GNNs, leading to trade-offs in accuracy depending on the specific task and available data [61]. The choice between these approaches depends on the nature of the available data (structured coordinates vs. text descriptions), the importance of interpretability vs. raw predictive power, and the specific property being predicted. LLM-based methods provide a complementary pathway, enriching the toolkit for understanding and predicting electrocatalyst behavior.

3.3 Hypothesis Generation and Synthesis Planning

Beyond analyzing existing knowledge, LLMs are increasingly being explored for their generative capabilities—their potential to propose novel materials, structures, or experimental plans, thereby acting as creative partners in the scientific discovery process [10,27,48–50,58,63]. This represents a significant shift towards using AI not just for prediction but also for exploration and innovation in electrocatalyst design.

One application is the generation of novel material compositions or structures. Inspired by the success of LLMs in generating text and code, researchers are training them on representations of materials to suggest new candidates. CatGPT, for example, is a GPT-based model trained on millions of catalyst structures (represented as tokenized strings of lattice parameters, atomic symbols, and coordinates) from the Open Catalyst database [64,65]. It can autoregressively generate string representations of new inorganic catalyst structures, including surface and adsorbate atoms. Fine-tuning CatGPT on a smaller dataset specific to the two-electron oxygen reduction reaction (2e-ORR) enabled the discovery of five novel and promising 2e-ORR catalyst candidates, demonstrating the potential for targeted discovery. Ensuring the chemical and structural validity of generated structures remains a challenge, requiring specialized validity metrics or post-generation filtering [64,65]. Other work explores generating crystal structures using LLMs trained on crystallographic information file (CIF) formats [66].

Frameworks are also being developed to leverage LLMs for generating scientific hypotheses. MOOSE-Chem is a multi-agent system that uses LLMs to tackle the complex task of hypothesis generation in chemistry [49]. MOOSE-Chem is a novel multi-agent LLM system developed to simulate the multi-step process of hypothesis generation in chemistry. It decomposes this process into document retrieval, knowledge synthesis, and hypothesis formulation. Notably, MOOSE-Chem was shown to independently rediscover hypotheses similar to those in high-impact 2024 papers, based on literature only available up to 2023. This demonstrates its potential to uncover latent connections in existing literature. However, its effectiveness heavily depends on the quality of the retrieval module and the diversity of the source documents. It breaks the process down into stages: retrieving relevant "inspiration" papers from literature based on a background question, synthesizing novel hypotheses by combining the background and inspirations and evaluating the quality of the generated hypotheses. Using LLMs trained on data up to 2023, MOOSE-Chem was able to rediscover hypotheses from high-impact chemistry papers published in 2024 with high similarity, showcasing the potential of LLMs to autonomously generate scientifically valid and novel ideas. Another approach, LLM-Feynman, combines LLM reasoning with optimization techniques to discover interpretable scientific formulae and theories from data, successfully rediscovering physics formulae and deriving accurate formulae for materials properties like synthesizability and ionic conductivity [50].

LLMs are also proving adept at synthesis planning and optimization, moving beyond simply extracting procedures to actively proposing or refining them. The ChemCrow agent, powered by GPT-4 and equipped with 18 expert-designed tools, demonstrated the ability to autonomously plan and conceptually execute the synthesis of organic molecules, including organocatalysts [10]. In the realm of inorganic nanomate-rials, a framework was developed using fine-tuned open-source LLMs for the synthesis of quantum dots (QDs) [59,67,68]. A dedicated QD synthesis planner framework was recently introduced, combining a fine-tuned LLM for protocol generation with a property prediction module. This system takes as input the desired properties (e.g., emission wavelength, particle size) and a masked base protocol, and outputs optimized synthetic recipes. The generated protocols underwent computational validation, expert review,

and experimental testing. Impressively, 3 of 6 proposed protocols succeeded in advancing the multiobjective Pareto front. This study illustrated the feasibility of integrating LLMs into closed-loop design cycles, though challenges remain in ensuring generalizability and robustness across materials domains. This system integrates a protocol generation model (taking target properties and a masked reference protocol as input) and a property prediction model. Generated protocols are validated computationally, assessed for novelty, evaluated by humans, and finally tested experimentally. This LLM-driven approach successfully generated QD synthesis protocols that improved target properties and updated the Pareto front for multi-objective optimization [68]. LLMs have also been used to suggest element libraries for exploring HEAs electrocatalysts for the ORR, providing a starting point for high-throughput experimental screening [34]. Furthermore, LLMs can predict suitable precursors for synthesizing target inorganic materials based on patterns learned from the literature [59] and assist in designing continuous flow microreactor systems [69].

Finally, LLMs can provide design recommendations or suggest specific experimental strategies. The CataLM model, for instance, was evaluated on a control method recommendation task for electrocatalytic materials, leveraging its fine-tuned knowledge base [44]. In a study on electrochemical C-H oxidation, LLMs were prompted to generate Python code that iteratively optimized reaction conditions to improve yields, demonstrating a collaborative human-AI approach to synthesis optimization (Fig. 5) [51]. These examples highlight the potential of LLMs to not only generate static plans but also participate in dynamic optimization loops.



Figure 5: (Continued)



Figure 5: (**A**) Design and assembly of the 24-well electrochemical platform and the schematic overview of the electrochemical $C(sp^3)$ –H oxidation process using the electrocatalyst; (**B**) Semantic literature analysis for reaction data mining using a language model with natural language prompts. Performance is evaluated by comparing the ground truth with LLM-assigned labels and examining the impact of prompt quality; (**C**) Overview of training data preparation and machine learning models used for predicting electrochemical C–H oxidation reactivity (Task 1) and selectivity (Task 2). Models with different architectures are evaluated for accuracy and AUC [51]

The application of LLMs in hypothesis generation and synthesis planning is still nascent but holds immense promise for accelerating the creative and planning aspects of electrocatalyst research. Key challenges include ensuring the scientific validity and synthesizability of generated outputs and effectively integrating these AI-driven suggestions with experimental validation and refinement.

3.4 Knowledge Synthesis and Integration

The exponential growth of scientific literature presents a significant challenge for researchers seeking to stay abreast of developments and synthesize existing knowledge [29]. LLMs, with their advanced natural language understanding and generation capabilities, offer powerful tools to manage this information deluge, integrate findings from disparate sources, and present synthesized knowledge in accessible formats [52].

One direct application is automated literature review and summarization. LLMs can process large numbers of research papers on a specific topic (e.g., propane dehydrogenation catalysts, CO₂ reduction catalysts) and generate comprehensive summaries or structured reviews [29,70]. These automated systems can significantly reduce the cognitive load on researchers and accelerate the process of understanding the state of the art. However, ensuring the factual accuracy and proper citation integrity of these generated reviews is critical. Multi-tier quality control strategies, potentially involving validation against multiple LLMs or expert verification, are necessary to mitigate the risk of hallucinations—the generation of plausible but false information—which is a known issue with current LLMs, especially in specialized domains. Studies have shown that with careful quality control, hallucination risks in generated reviews can be reduced significantly [29].

LLMs can also facilitate knowledge integration through knowledge graphs (KGs). KGs provide a structured way to represent entities (like materials, properties, and reactions) and the relationships between them [40]. LLMs can be employed in the construction of these KGs by automatically extracting entities and relationships from text. Furthermore, LLMs can interact with existing KGs, allowing researchers to query this structured knowledge base using natural language [25]. The synergy between the semantic understanding

of LLMs and the structured representation of KGs is particularly powerful. For example, combining word embeddings (capturing textual context) with graph embeddings (capturing relational structure from a KG of Cu-based CO₂RR catalysts) led to improved prediction of FE, demonstrating how LLMs can leverage both unstructured and structured knowledge sources.

By processing vast amounts of extracted data or interacting with KGs, LLMs can assist in trend analysis and insight generation. They can help identify historical developments in catalyst research, pinpoint emerging materials or synthesis techniques, and uncover correlations between different factors influencing catalytic performance (e.g., relationships between catalyst composition, regulation methods, and product selectivity in CO₂RR) [40,70]. This ability to synthesize information across numerous studies can potentially reveal hidden connections or patterns that might be missed through manual review [19].

The development of intelligent Q&A systems tailored for electrocatalysis is another promising direction. These systems, often built using LLMs fine-tuned on domain literature or grounded using RAG techniques with extracted data or KGs, allow researchers to ask specific questions and receive informative answers supported by evidence from the literature [25,37]. An example is the development of a data-grounded chatbot for answering questions about MOF synthesis procedures [37].

In essence, LLMs are emerging as crucial tools for knowledge management and synthesis in the rapidly expanding field of electrocatalysis. By automating literature reviews, facilitating the construction and use of knowledge graphs, enabling trend analysis, and powering intelligent Q&A systems, they help researchers navigate, understand, and build upon the collective knowledge embedded within scientific publications, ultimately fostering a more integrated and accelerated discovery process.

4 Applications, Performance Analysis, and Comparative Insights

4.1 Case Studies across Electrocatalytic Reactions and Materials

The versatility of LLMs is reflected in their application across the major electrocatalytic reactions critical for energy and environmental technologies. For the HER, which is fundamental to water splitting for hydrogen fuel production, LLMs and related NLP/ML techniques are being used to accelerate the discovery of efficient, low-cost catalysts, often aiming to replace expensive platinum-group metals [15,71]. ML frameworks have been developed to screen large numbers of candidate alloys, identifying promising high-performance materials like AgPd, whose potential was subsequently verified experimentally and computationally under realistic conditions (Fig. 6) [72]. Research specifically targets low-dimensional materials like nanoparticles, nanotubes, and nanosheets, leveraging ML to predict their HER performance based on various descriptors [15]. Text mining combined with Word2Vec embeddings and Pareto optimization has been applied to predict HER-active candidate compositions based on their textual similarity to relevant properties like conductivity [14]. LLMs are also envisioned to provide direct design guidance for HER catalysts [42]. High-throughput experimental methods combined with data-driven strategies, potentially guided by LLM-suggested element libraries, are accelerating the discovery of HER-active HEAs [73].

The OER, the typically sluggish anodic counterpart to HER in water splitting, is another key target. LLMs are being applied for predictive analytics, for example, in the context of (oxy)hydroxide-based OER catalysts [36]. While details are emerging, this suggests LLMs are used for tasks like extracting performance data from literature or predicting activity based on compositional or textual features for these materials. ML, often coupled with high-throughput experiments or simulations, is actively used to discover and optimize OER catalysts, including perovskite oxides, where active learning approaches have identified compositions with exceptionally low overpotentials. Text mining and embedding-based approaches, similar to those used for HER, are also applied to predict promising OER candidate compositions [14]. LLMs are expected to

play a role in providing design guidance for OER catalysts as well [42]. Interpretable ML models are being developed to unify activity prediction across multiple reactions, including OER, using intrinsic material properties [39].



Figure 6: An ML framework for high-throughput screening of electrocatalysts. In the left of the section "Constructing Adsorption Database": Adsorption sites for binary alloys, including ontop, bridge, and hollow sites, denoted by black star, red "+", and blue "×", respectively [72]

The ORR is crucial for fuel cells and metal-air batteries. LLM-based generative models like CatGPT have been fine-tuned specifically for discovering catalysts for the selective two-electron ORR pathway, which produces H_2O_2 , a valuable chemical [64,65]. The model successfully generated novel candidate structures validated by further analysis. In another approach, an LLM provided an initial element library to guide the high-throughput experimental discovery of Pt-based quinary HEAs for ORR [34]. This involved microscale precursor printing and rapid screening with scanning electrochemical cell microscopy (SECCM), demonstrating a powerful synergy between LLM guidance and automated experimentation. Text mining and embedding-based methods are also used for predicting ORR candidates [14], and LLMs are anticipated to offer design guidance [42]. ML models are also being developed to optimize HEA compositions for ORR activity [35].

The CO₂RR aims to convert CO₂ into valuable fuels and feedstocks, mitigating greenhouse gas emissions. This area has seen significant LLM application, particularly in knowledge extraction and synthesis. LLM-enhanced methods have been used to create large corpora detailing CO₂RR electrocatalysts (beyond just Cu-based systems) and their synthesis procedures by extracting information like materials, products, Faradaic efficiencies (FE), and experimental conditions from thousands of papers [54]. NLP tools have been used to analyze trends in non-Cu catalysts from literature, identifying emerging materials like perovskites

and bismuth oxyhalides [70]. A notable study constructed a knowledge graph specifically for Cu-based CO_2RR catalysts using a SciBERT-based framework [40]. This KG not only visualized development trends but was also used, by combining word and graph embeddings, to predict the FE for specific products, showcasing the integration of structured and unstructured knowledge. LLMs are also expected to provide design guidance for CO_2RR catalysts [42].

Beyond these specific reactions, LLMs are being applied to broader tasks in catalyst research relevant to electrocatalysis. This includes text mining and prediction for MOF synthesis using ChatGPT, LLMdriven synthesis planning for QDs which can have electrocatalytic applications [68], IE for single-atom heterogeneous catalysts [58], and assisting in the exploration of electrochemical C-H oxidation reactions through literature mining and code generation for ML model training [51]. These case studies illustrate the diverse ways LLMs are beginning to impact the field, from large-scale data aggregation to targeted material prediction of LLMs across major electrocatalytic reactions such as HER, OER, ORR, and CO₂RR, organizing them by task type (e.g., property prediction, synthesis planning, knowledge extraction) and highlighting specific use cases and insights. For example, in the case of CO₂RR, LLMs have been utilized for corpus creation, trend analysis, and FE prediction using hybrid embedding techniques. For HER and OER, LLM-guided screening has facilitated the identification of promising alloy compositions. This table crystallizes how LLMs function as modular tools across the catalyst discovery landscape, tailored to the nuances of different electrochemical reactions.

Reaction	LLM task	Specific examples/Refs.	Key insights/Findings for reaction
HER	Property	ML framework for alloy	Acceleration of alloy
	prediction/Design	discovery (AgPd) [72]; ML for	screening; Prediction
		low-dim catalysts [15]; Text	based on text similarity;
		mining/Word2Vec	Guidance for catalyst
		prediction [14]; LLM design	design
		guidance [42]; HEA discovery	
		w/HT expts [73]	
OER	Property	LLM predictive analytics for	Prediction for specific
	prediction/Design	(oxy)hydroxides [36]; ML/AI	material classes;
		for perovskites [36]; Text	Text-based candidate
		mining/Word2Vec	screening; Design
		prediction [14]; LLM design	guidance
		guidance [42]; Interpretable ML	
		for activity [39]	
	Structure	CatGPT fine-tuned for 2e-ORR	Discovery of novel
ORR	generation/Design	catalyst discovery [65]	2e-ORR candidates
	Element	LLM generating element library	LLM-guided
	selection/Design	for Pt-HEA discovery + HT	high-throughput
		expts [34]	discovery workflow

Table 2: Overview of LLM applications in specific electrocatalytic reactions

(Continued)

Reaction	LLM task	Specific examples/Refs.	Key insights/Findings for reaction
	Property	Text mining/Word2Vec	Text-based candidate
	prediction/Design	prediction [14]; LLM design	screening; Design
		guidance [42]; ML optimization	guidance
	IE/Commune enaction	IOI FIEAS [55]	I ango poplo data
	IE/Corpus creation	(general CO PD) [54]: NI D	Large-scale data
$CO_2 KR$		(general $CO_2 RR$) [54]; NLP	aggregation; frend
		review of non-Cu catalysts [70]	analysis beyond Cu
	KG Construction/FE	KG for Cu-based catalysts	Structured knowledge
	Prediction	(SciBERT); FE prediction via	representation;
		word + graph embeddings [40]	Prediction integrating
			semantics & structure
	Design guidance	LLM design guidance [42]	Guidance for catalyst
			design
	Synthesis planning	LLM for QD synthesis [68];	Optimization of
General/Other	, , ,	ChatGPT for MOF	nanomaterial synthesis
		synthesis [37]	protocols
	IE	LLM for MOF synthesis [37];	Data extraction for
		LLM for single-atom	specific catalyst
		catalysts [58]; LLM for C-H	types/reactions
		oxidation lit. mining [51]	
	Property prediction	CatBERTa for adsorption	Text-based prediction of
		energy [61]	fundamental catalytic
			properties

Table 2 (continued)

4.2 Comparative Analysis of LLM Approaches

As LLMs become more integrated into electrocatalysis research, understanding the nuances between different models, training strategies, and application frameworks is crucial for selecting and developing effective tools.

Regarding model architectures, the field utilizes both BERT-based encoders (like SciBERT, RoBERTa used in CatBERTa, MatBERT) primarily for understanding and classification tasks, and GPT-style decoders or encoder-decoders for generative tasks like text generation, hypothesis generation, or synthesis planning [40,44,46,61,68]. BERT-based models excel at extracting semantic meaning and have shown strong performance in NER and property prediction from text [40,60,61]. GPT-based models, with their strong generative capabilities, are increasingly used for proposing synthesis routes, generating novel structures, or acting as conversational agents [37,49,65,68]. The choice of architecture often depends on the primary goal, whether it is analyzing existing text or generating new information.

Training strategies significantly impact performance. While general-purpose LLMs like GPT-4 can perform reasonably well on some tasks using sophisticated prompt engineering [37,46], studies consistently show that domain-specific pre-training or fine-tuning yields substantial improvements for specialized scientific tasks [10,44,54,60]. Models like MatBERT (pre-trained on materials science literature) [60], CatBERTa

(fine-tuned RoBERTa for catalyst energy prediction) [61], and CataLM (Vicuna fine-tuned on electrocatalysis literature and expert data) [44] demonstrate enhanced understanding of domain terminology and concepts, leading to higher accuracy in tasks like NER, property prediction, and recommendation [44,60,61]. Parameter-Efficient Fine-Tuning (PEFT) techniques are also being adopted to adapt large models to specific tasks more efficiently, as seen in the QD synthesis planning framework [68]. The trade-off lies between the versatility of large general models and the specialized accuracy of fine-tuned models, with the latter often preferred for demanding scientific applications where domain knowledge is critical.

The type of input data processed by the LLM also differentiates approaches. Many applications focus purely on textual input, leveraging LLMs' core strength in NLP [60,61]. CatBERTa predicts energy from textual descriptions [61], and MOOSE-Chem generates hypotheses from background questions and literature text [49]. Other methods integrate structured data or embeddings. The Cu-CO₂RR KG study combined word embeddings from text with graph embeddings from the structured KG for FE prediction [40]. Word2Vec embeddings derived from abstracts were used as numerical inputs for Pareto optimization [14]. The future direction clearly points towards multimodal inputs, integrating text with figures, tables, and potentially experimental data streams, although this remains a significant developmental challenge [27,42].

Finally, LLMs are being applied both as standalone tools and as components within larger integrated workflows. Standalone applications might involve using an LLM directly for Q&A, summarization, or prediction based on a prompt [37,61]. Integrated approaches are becoming more common and powerful. Examples include using LLM-based IE to populate databases that feed into downstream ML models [56], employing LLMs to generate code for ML model training or optimization [51], combining LLMs with knowledge graphs [40], and embedding LLMs within autonomous laboratory frameworks to guide experiments [34]. These integrated systems leverage the LLM's language and reasoning capabilities while connecting them to other computational tools or physical experiments, amplifying their impact. The trend suggests a move towards more sophisticated, integrated systems where LLMs act as orchestrators or intelligent interfaces within broader scientific discovery pipelines. Table 3 presents a systematic comparison of various LLM frameworks applied to IE tasks within electrocatalysis. It categorizes each model by the specific task (e.g., band gap extraction, synthesis parameter identification), the source and size of input data, and the resulting output format. Additionally, the table evaluates each model's effectiveness and limitations based on published benchmarks and practical deployment. For instance, GPT-4 demonstrates strong performance in general extraction tasks but suffers from hallucination and high computational costs. In contrast, domain-specific models like CataLM and SciBERT-BiLSTM-CRF offer improved accuracy in materials entity recognition within defined contexts (e.g., CO_2RR), albeit at the cost of requiring domain-specific training data. This comparative analysis highlights the trade-offs between generalizability and domain accuracy in current LLM-powered IE pipelines.

Model/Framework	Task	Input data (Source, Size)	Output format	Comparative evaluation: effectiveness, efficiency, and limitations	Refs.
GPT-4	Band gap extraction	415 random materials science articles (manual eval)	Band gap values, materials	Effectiveness: 88% accuracy. Efficiency: High GPU demand (e.g., A100). Limitations: Prone to hallucination, general-domain model.	[46]

Table 3: Overview of LLMs applied to electrocatalyst information extraction

(Continued)

Model/Framewor	k Task	Input data (Source, Size)	Output format	Comparative evaluation: effectiveness, efficiency, and limitations	Refs.
ChatGPT (GPT-3.5/4)	MOF synthesis extraction	Peer-reviewed MOF articles (~800 MOFs)	26k+ synthesis parameters (structured)	Effectiveness: F1 90%–99%. Efficiency: Fast inference with engineered prompts. Limitations: Context-length sensitivity, depends on prompt clority	[37]
SciBERT + BiLSTM- CRF/BERT- BiLSTM-CRF	CO ₂ RR entity recognition	835 publications (benchmark); 372 full texts (extended)	Entities (Material, Method, Product, FE, etc.)	Effectiveness: Micro-F1 ~82%. Efficiency: Moderate training cost. Limitations: Requires annotated corpora, limited to CO ₂ RR context.	[56]
LLMs (General) + NLP Techniques	CO ₂ RR corpus creation & synthesis extraction	5941 documents (metadata); 2776 full texts	Benchmark corpus (7k records); Extended corpora (77k + 30k records); Action sequences	Effectiveness: Comprehensive corpus created. Efficiency: Costly pretraining + fine-tuning. Limitations: Requires large, domain-specific data.	[54]
ChatExtract (Conversa- tional LLM, e.g., GPT-4)	General materials data extraction	Research papers (test sets mentioned)	Extracted data points	Effectiveness: ~90% precision/recall. Efficiency: Lightweight conversational method. Limitations: Hallucination possible, context-sensitive	[36]
NLP Tools (General)	CO2RR literature review (non-Cu)	7292 published articles	Trends, emerging materials (perovskites, Bi-oxyhalides), common elements, electrolytes	Effectiveness: Qualitative trend mapping. Efficiency: Text-only mining; low compute need. Limitations: Lacks structured extraction.	[70]
CataLM (Fine-tuned Vicuna-13B)	Entity extraction (Electrocataly- sis)	Domain literature, expert annotations	Entities (Material, Method, Product, FE, etc.)	Effectiveness: High domain-specific accuracy. Efficiency: PEFT applied for tuning. Limitations: Performance drops outside domain.	[44]
Transformer (ACE)	Information extraction (SACs)	Literature	Structured data	Effectiveness: Focused and relevant SAC data. Efficiency: Narrow application. Limitations: Unclear generalizability, lacks metric benchmarks.	[58]

4.3 Performance Evaluation and Benchmarking

Assessing the performance of LLMs in the specialized domain of electrocatalysis requires appropriate metrics and rigorous evaluation, although standardized benchmarks are still largely lacking [16,22]. The metrics used vary depending on the specific task being performed. For IE tasks like NER and relation extraction, standard NLP metrics such as Precision, Recall, and F1-score are commonly employed [36,37]. These metrics quantify the accuracy of identifying entities and relationships compared to a ground truth

(often manually annotated data). For instance, the ChatGPT-based MOF synthesis extraction achieved F1 scores of 90%–99% using careful prompt engineering, while the BERT-BiLSTM-CRF model for CO_2RR entity recognition reached micro-average F1 scores of around 82% [56]. The ChatExtract method reported precision and recall close to 90% for materials data extraction.

In property prediction, common regression metrics are used when predicting continuous values like adsorption energy or FE. MAE is frequently reported, indicating the average absolute difference between predicted and true values [60,61,72]. CatBERTa reported MAEs between 0.35 and 0.75 eV for adsorption energy prediction, depending on the dataset and input features [61]. The R² is also used to measure the proportion of variance explained by the model [50]. For classification tasks (e.g., predicting synthesizability or classifying materials based on properties), accuracy and metrics like the Matthews correlation coefficient are used [60].

Evaluating generative and planning tasks requires different approaches. For hypothesis generation, metrics might include similarity scores to known valid hypotheses or expert evaluation of novelty and feasibility [49]. For structure generation, validity checks (e.g., detecting overlapping atoms, and ensuring chemical sensibility) are crucial, alongside assessing the novelty and predicted properties of the generated structures [64]. For synthesis planning (e.g., LLM-driven QD synthesis), success can be measured by the rate at which generated protocols lead to successful experiments, improve target properties, or advance the Pareto front in multi-objective optimization [68]. The QD synthesis work reported that 3 out of 6 LLM-generated protocols updated the Pareto front.

Beyond accuracy metrics, efficiency is also a key evaluation criterion. This can involve measuring the reduction in time required for tasks like literature review (e.g., seconds per article for automated review generation [29]) or the potential reduction in computational or experimental cost achieved by using LLM predictions to guide efforts [56].

Despite these reported successes, a significant benchmarking gap exists. There is a lack of standardized, publicly available benchmark datasets and evaluation protocols specifically designed for testing LLMs on various tasks within electrocatalysis or even the broader field of materials science [16,22]. This makes it challenging to directly compare the performance of different models and methodologies developed in separate studies. Establishing such benchmarks would be crucial for driving progress and ensuring rigorous assessment of new LLM approaches. Current evaluations often rely on specific internal datasets or comparisons against limited baselines, highlighting the need for community-wide efforts in developing standardized evaluation frameworks. Table 4 summarizes the predictive performance of different LLM-based frameworks in estimating key electrocatalytic properties such as adsorption energy and FE. The models are organized by input type (e.g., textual descriptions, word or graph embeddings), predicted properties, target catalyst systems, and quantitative performance metrics (e.g., MAE, accuracy). For example, CatBERTa-trained on structured textual inputs like composition and surface features-achieved an MAE of 0.35-0.75 eV for adsorption energy predictions, making it a compelling alternative to more computationally demanding Graph Neural Networks. Other models, such as those using hybrid word-graph embeddings, demonstrated strong performance in FE prediction tasks for Cu-based CO₂RR catalysts. The table emphasizes how LLMs can complement or even rival traditional methods by enabling property prediction from semantically rich, yet structurally limited, inputs. These insights are particularly relevant for scenarios where experimental or atomic-scale data is sparse or unavailable. Table 5 highlights key frameworks and models that harness the generative and planning capabilities of LLMs within the realm of electrocatalyst research. It details each framework's core task (e.g., hypothesis generation, structure proposal, synthesis optimization), material focus, the LLM's specific role (e.g., retriever, generator, predictor), and outcomes or validation methods. For instance, MOOSE-Chem uses a multi-agent architecture for autonomous hypothesis formation in chemistry, while CatGPT excels at proposing valid catalyst structures for 2e-ORR through fine-tuned autoregressive generation. Other systems like the LLM-Feynman framework demonstrate symbolic regression abilities, rediscovering known physical laws, while QD-focused models successfully improved experimental synthesis outcomes using Pareto optimization. This table illustrates the growing sophistication of LLM-enabled systems capable of creativity, planning, and real-world lab guidance, bridging the gap between literature mining and experimental realization.

Model/Framework	Input type	Predicted property	Material system/Reaction	Performance metric/Value	Comparative evaluation: effectiveness, efficiency, and limitations	Refs.
CatBERTa (RoBERTa- based)	Textual description (Composition, Structure features)	Adsorption energy	General catalysts (OC20 dataset)	MAE: 0.75 eV (100k data); MAE: 0.35 eV (high-accuracy subset)	Effectiveness: MAE 0.35–0.75 eV. Efficiency: Requires moderate fine-tuning. Limitations: Less precise than GNNs for local atomic interactions.	[61]
Word2Vec + Pareto Optimization	Word embeddings (from Abstracts)	Similarity to 'Conductiv- ity'/'Dielectric' (proxy for activity)	Candidate Compositions (e.g., AgPdPt, AgPdRu)/HER, OER, ORR	Predicted high-performing compositions matched experimental trends	Effectiveness: High match with experiment. Efficiency: Lightweight; corpus-based. Limitations: Relies on textual semantic similarity, lacks explicit structural context.	[14]
LLM (SciBERT) + KG Embeddings + DL	Word embeddings + Graph Embeddings	FE	Cu-based Catalysts/CO2RR	Demonstrated FE prediction capability	Effectiveness: Good FE prediction. Efficiency: Requires graph construction. Limitations: Dependent on quality of both text and KG data.	[40]
Transformer (MatBERT)	Textual Description (Robocrystallog- rapher)	Energy above hull, Band gap, SLME, Spin-orbit spillage, Magnetic moment (Classification)	General inorganic materials	Accuracy > GNNs (ALIGNN) for 4/5 properties; High accuracy even with small datasets	Effectiveness: High classification accuracy. Efficiency: Pretrained model with lightweight fine-tuning. Limitations: Performance depends on Robocrystallographer input quality.	[60]

 Table 4: Performance comparison of LLMs for electrocatalyst property prediction

Table 5: LLM-driven	hypothesis	generation and s	ynthesis	planning exam	ples
			1		

Framework/Model	Task	Material focus	LLM role	Key outcome/Validation	Comparative evaluation: effectiveness, efficiency, and limitations	Refs.
MOOSE-Chem	Hypothesis generation	General chemistry	Retrieval, synthesis, ranking (Multi-agent)	Rediscovered novel hypotheses from 2024 papers with high similarity	Effectiveness: High novelty similarity. Efficiency: Multi-agent but scalable. Limitations: Performance depends on retriever quality.	[49]

(Continued)

Table 5 (continued)						
Framework/Model	Task	Material focus	LLM role	Key outcome/Validation	Comparative evaluation: effectiveness, efficiency, and limitations	Refs.
CatGPT (GPT-based)	Structure generation	Inorganic catalysts (OC20)/2e-ORR Catalysts	Generator	Generated valid catalyst structures; Discovered 5 novel 2e-ORR candidates after fine-tuning	Effectiveness: Discovered 5 new candidates. Efficiency: Heavy compute during generation. Limitations: Validity filtering needed post-generation	[65]
LLM-Feynman	Formula/Theory discovery	Physics, materials science (Synthesizability, ionic conductivity, bandgap)	Symbolic regression, interpreter	Rediscovered >90% physics formulae; Derived interpretable formulae for materials properties (Acc > 90%, R^2 > 0.8)	Effectiveness: >90% formula rediscovery. Efficiency: Symbolic regression efficient. Limitations: May not scale to highly complex systems.	[50]
LLM-driven framework (Fine-tuned Open LLM + PEFT)	Synthesis plan- ning/Optimization	Quantum Dots (QDs)	Generator, Predictor	Generated 6 protocols; 3 updated Pareto front; All improved ≥ 1 property; Validated experimentally	Effectiveness: All generated protocols improved some targets. Efficiency: PEFT used. Limitations: Dependent on accurate property predictor.	[59]
LLM (unspecified)	Element selection	Pt-based high-entropy alloys (HEAs)/ORR	Generator (Element library)	Provided element library guiding high-throughput experimental screening (SECCM)	Effectiveness: Enabled targeted HT screening. Efficiency: Fast suggestion generation. Limitations: No accuracy benchmark reported.	[34]
LLM (unspecified) + Code Generation	Synthesis optimization	Electrochemical C-H Oxidation	Code generator (for ML optimization)	Iteratively improved reaction yields based on prompts; Optimized conditions for 8 drug-like substrates	Effectiveness: Improved multi-objective outcomes. Efficiency: Code auto-generation fast. Limitations: Prompt dependency and generalizability.	[51]
ChemCrow (GPT-4 + Tools)	Synthesis planning	Organic molecules (incl. Organocatalysts)	Planner, tool-user (Agent)	Autonomously planned and executed syntheses	Effectiveness: Demonstrated successful planning. Efficiency: Combined with external tools. Limitations: Hallucination and chaining logic errors	[10]
CataLM (Fine-tuned Vicuna-13B)	Control method recommendation	Electrocatalytic Materials	Recommender	Validated on recommendation task using domain knowledge	Effectiveness: Relevant method suggestions. Efficiency: Fast inference. Limitations: Domain-specific tuning required.	[44]

5 Challenges, Limitations, and Future Outlook

While the application of LLMs in electrocatalyst research holds significant promise, the field faces substantial challenges and limitations that must be addressed to realize its full potential. Concurrently, exciting future directions are emerging, pointing towards more powerful and integrated AI-driven discovery workflows.

5.1 Current Challenges and Limitations

Several key hurdles currently impede the widespread and reliable application of LLMs in electrocatalyst design. A fundamental issue lies with data. Despite the vastness of scientific literature, accessing high-quality, comprehensive, and standardized data remains difficult [10,16,22]. LLM training requires large datasets, but electrocatalysis data can be sparse for specific materials or reactions, heterogeneous due to varying experimental conditions and reporting standards, and potentially biased due to the tendency to underreport negative results [13,22,27]. Extracting complete information is further complicated by knowledge fragmented across multiple publications and supplementary materials, which current LLMs struggle to synthesize effectively [27].

The interpretability and explainability of LLMs pose another significant challenge. Many state-of-the-art models function as "black boxes", making it difficult to understand why they make a particular prediction or suggestion [22]. This lack of transparency hinders scientific understanding, trust in the model's output, and the ability to extract generalizable design principles [10]. While techniques like attention analysis or post-hoc explanations offer some insight [61], achieving true mechanistic understanding from LLM outputs remains an open research area.

Hallucinations and factual inaccuracy represent a critical barrier to the reliable use of LLMs in science [27,46,50,54]. LLMs can generate text that sounds scientifically plausible but is factually incorrect, unsubstantiated, or physically inconsistent. This risk is potentially amplified in specialized domains like electrocatalysis where the training data might be less comprehensive compared to general text [29]. Rigorous validation and mitigation strategies, such as RAG, careful prompting, and expert verification, are essential but add complexity to the workflow [29,36].

Current LLMs also exhibit limitations in domain knowledge grounding and reasoning [27,50]. While they possess vast general knowledge, their understanding of fundamental materials science and electrochemistry principles can be shallow. They struggle with complex numerical reasoning, unit conversions, understanding intricate chemical notations (e.g., varied formulae, crystallographic information like CIF files), and applying core concepts like crystal symmetry or reaction stoichiometry correctly [27]. This limits their ability to perform deep, physically grounded reasoning. The predominantly text-based nature of most current LLMs restricts their ability to process multimodal data. Scientific publications in materials science heavily rely on figures (micrographs, diffraction patterns, performance plots), complex tables, and chemical structure diagrams to convey crucial information. LLMs that cannot interpret these visual or tabular formats miss a significant portion of the available knowledge.

Finally, even with advanced AI predictions, the experimental validation bottleneck persists [10,27]. Any catalyst candidate or synthesis plan proposed by an LLM must ultimately be tested in the laboratory, a process that remains resource-intensive and time-consuming. LLMs primarily accelerate the *in silico* stages of discovery and design.

5.2 Future Research Directions

Addressing the current limitations and harnessing the full potential of LLMs in electrocatalysis necessitates focused research efforts along several key directions.

A critical need is the development of specialized Materials Science LLMs (MatSci-LLMs). This involves moving beyond general-purpose models towards architectures pre-trained or extensively fine-tuned on vast corpora of materials science literature, textbooks, and databases. Such models would possess deeper domain knowledge, better understand specialized terminology and notations (including chemical formulas and

crystallographic data), and exhibit improved reasoning capabilities grounded in materials science principles. Examples like MatSciBERT, CataLM, and CatBERTa represent early steps in this direction [44,61].

The development of multimodal LLMs is arguably one of the most crucial future directions. Models capable of seamlessly integrating and reasoning over information from text, tables, figures (plots, images, schematics), chemical structures, and potentially experimental data streams (e.g., spectra) will unlock a much larger fraction of scientific knowledge and enable more holistic analysis [27,42].

Improving reasoning capabilities and grounding LLM outputs in fundamental physical and chemical laws is essential for generating scientifically valid and reliable predictions or hypotheses. Integrating LLMs with physics-informed AI principles, where physical constraints or knowledge are incorporated into the model architecture or training process, holds significant promise [10,17,22].

Enhancing interpretability and explainability (XAI) remains paramount for building trust and extracting scientific insights [22,60,74]. Future work should focus on developing robust XAI techniques tailored for LLMs in scientific domains, allowing researchers to understand the basis of model predictions and identify potential failure modes.

Continued research into mitigating hallucinations is vital for scientific applications where factual accuracy is non-negotiable. This includes refining techniques like RAG, developing better-prompting strategies, improving fine-tuning methods focused on factuality, incorporating self-evaluation mechanisms within LLMs, and utilizing multi-agent frameworks for cross-validation [49].

The community needs to establish standardized datasets and benchmarks for evaluating LLMs on various electrocatalysis-related tasks [16,22]. This will enable objective comparison of different models and methodologies, track progress more effectively, and identify areas needing further improvement.

A highly promising future direction is the integration of LLMs with autonomous experimentation platforms or Self-Driving Laboratories (SDLs) [75]. In such closed-loop systems, LLMs could analyze previous results, consult literature knowledge, propose the next set of experiments, generate the necessary code to control robotic hardware, interpret incoming data, and iteratively refine the search for optimal catalysts or synthesis conditions. This synergy between AI-driven decision-making and automated execution has the potential to dramatically accelerate the pace of discovery.

6 Conclusion

The integration of LLMs into the field of electrocatalysis represents a nascent but rapidly advancing frontier with the potential to reshape materials discovery and design. This review has highlighted the diverse methodologies being employed, spanning automated information extraction from the vast scientific literature, novel approaches to property prediction based on textual data and embeddings, the generation of hypotheses for new materials and synthesis routes, and the synthesis of knowledge scattered across countless publications. Case studies across critical reactions like HER, OER, ORR, and CO2RR demonstrate tangible progress, with LLMs contributing to the creation of valuable datasets, the prediction of catalytic performance, and even the suggestion of novel catalyst candidates and optimized synthesis protocols. Models specifically adapted to the materials science domain, such as CatBERTa and CataLM, alongside innovative frameworks like MOOSE-Chem and LLM-driven synthesis planners, showcase the growing sophistication of these tools. However, significant challenges remain. Issues surrounding data availability and quality, the inherent "black-box" nature and potential for hallucinations in LLMs, limitations in deep scientific reasoning and multimodal data processing, and the persistent need for experimental validation must be rigorously addressed. Overcoming these hurdles will require concerted efforts in developing domain-specific and multimodal models, enhancing interpretability and factual grounding, establishing standardized

benchmarks, and fostering collaborative research practices. The fusion of LLMs with physics-informed AI, their incorporation into autonomous experimental workflows within self-driving laboratories, and their role as sophisticated collaborators alongside human researchers promise to significantly accelerate the pace of innovation. By effectively harnessing the ability of LLMs to process, synthesize, and generate knowledge from the ever-expanding body of scientific literature, we can anticipate a future where the rational design and discovery of advanced micro/nano electrocatalyst materials—crucial components for a sustainable energy landscape—is achieved with unprecedented speed and efficiency.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Hassan Karimi-Maleh and Li Fu; methodology, Ying Shen and Shichao Zhao; software, Ying Shen and Fei Chen; validation, Ying Shen, Shichao Zhao and Yanfei Lv; formal analysis, Ying Shen and Yanfei Lv; investigation, Ying Shen and Fei Chen; resources, Fei Chen and Li Fu; data curation, Ying Shen and Fei Chen; writing—original draft preparation, Ying Shen and Yanfei Lv; writing—review and editing, Hassan Karimi-Maleh; visualization, Ying Shen and Fei Chen; supervision, Li Fu and Hassan Karimi-Maleh; project administration, Li Fu and Hassan Karimi-Maleh. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The authors confirm that the data supporting the findings of this study are available within the article.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

- 1. Basyooni-M Kabatas MA. A comprehensive review on electrocatalytic applications of 2D metallenes. Nanomaterials. 2023;13(22):2966. doi:10.3390/nano13222966.
- Lucas FWS, Grim RG, Tacey SA, Downes CA, Hasse J, Roman AM, et al. Electrochemical routes for the valorization of biomass-derived feedstocks: from chemistry to application. ACS Energy Lett. 2021;67(11):1205–70. doi:10.1021/ acsenergylett.0c02692.
- 3. Tüysüz H. Alkaline water electrolysis for green hydrogen production. Acc Chem Res. 2024;57(4):558–67. doi:10. 1021/acs.accounts.3c00709.
- 4. Zhang S, Fan Q, Xia R, Meyer TJ. CO₂ reduction: from homogeneous to heterogeneous electrocatalysis. Acc Chem Res. 2020;53(1):255–64. doi:10.1021/acs.accounts.9b00496.
- 5. Braun M, Santana CS, Garcia AC, Andronescu C. From waste to value—glycerol electrooxidation for energy conversion and chemical production. Curr Opin Green Sustain Chem. 2023;41:100829. doi:10.1016/j.cogsc.2023. 100829.
- 6. Page JR, Manfredi Z, Bliznakov S, Valla JA. Recent progress in electrochemical upgrading of bio-oil model compounds and bio-oils to renewable fuels and platform chemicals. Materials. 2023;16(1):394. doi:10.3390/ma16010394.
- Kumar A, Bhayana S, Singh PK, Tripathi AD, Paul V, Balodi V, et al. Valorization of used cooking oil: challenges, current developments, life cycle assessment and future prospects. Discov Sustain. 2025;6(1):119. doi:10.1007/s43621-025-00905-7.
- 8. Garedew M, Lin F, Song B, DeWinter TM, Jackson JE, Saffron CM, et al. Greener routes to biomass waste valorization: lignin transformation through electrocatalysis for renewable chemicals and fuels production. ChemSusChem. 2020;13(17):4214–37. doi:10.1002/cssc.202000987.

- Ramírez Á., Muñoz-Morales M, Fernández-Morales FJ, Llanos J. Valorization of polluted biomass waste for manufacturing sustainable cathode materials for the production of hydrogen peroxide. Electrochim Acta. 2023;456:142383. doi:10.1016/j.electacta.2023.142383.
- 10. Xu Y, Wang H, Zhang W, Xie L, Chen Y, Salim F, et al. AI-empowered catalyst discovery: a survey from classical machine learning approaches to large language models. arXiv:2502.13626. 2025.
- 11. Xia Y, Campbell CT, Roldan Cuenya B, Mavrikakis M. Introduction: advanced materials and methods for catalysis and electrocatalysis by transition metals. Chem Rev. 2021;121(2):563–6. doi:10.1021/acs.chemrev.0c01269.
- 12. Huo W, Wang S, Dominguez-Gutierrez FJ, Ren K, Kurpaska Ł, Fang F, et al. High-entropy materials for electrocatalytic applications: a review of first principles modeling and simulations. Mater Res Lett. 2023;11(9):713–32. doi:10. 1080/21663831.2023.2224397.
- 13. Hu Y, Chen J, Wei Z, He Q, Zhao Y. Recent advances and applications of machine learning in electrocatalysis. J Mater Inf. 2023;3(3):18. doi:10.20517/jmi.2023.23.
- 14. Zhang L, Stricker M. Electrocatalyst discovery through text mining and multi-objective optimization. arXiv:2502.20860. 2025.
- 15. Zhao G, Huang Y, Li Y, Jena P, Zeng XC. Machine learning-assisted low-dimensional electrocatalysts design for energy conversion. Chem Rev. 2023;123(3):1398–454. doi:10.1021/acs.chemrev.2c00297.
- 16. Ding R, Chen J, Chen Y, Liu J, Bando Y, Wang X. Unlocking the potential: machine learning applications in electrocatalyst design for electrochemical hydrogen energy transformation. Chem Soc Rev. 2024;53(23):11390–461. doi:10.1039/d4cs00844h.
- 17. Wu H, Chen M, Cheng H, Yang T, Zeng M, Yang M. Interpretable physics-informed machine learning approaches to accelerate electrocatalyst development. J Mater Inf. 2025;5(2):15. doi:10.20517/jmi.2024.67.
- 18. Choudhary K, DeCost B, Chen C, Jain A, Tavazza F, Cohn R, et al. Recent advances and applications of deep learning methods in materials science. npj Comput Mater. 2022;8(1):59. doi:10.1038/s41524-022-00734-6.
- 19. Jiang X, Wang W, Tian S, Wang H, Lookman T, Su Y. Applications of natural language processing and large language models in materials discovery. npj Comput Mater. 2025;11(1):79. doi:10.1038/s41524-025-01554-0.
- 20. Shetty V, Shabari Shedthi B, Shashishekar C. Application and challenges of machine learning techniques in mining engineering and material science. J Mines Met Fuels. 2023:1989–2000. doi:10.18311/jmmf/2023/36099.
- 21. Yin G, Zhu H, Chen S, Li T, Wu C, Jia S, et al. Machine learning-assisted high-throughput screening for electrocatalytic hydrogen evolution reaction. Molecules. 2025;30(4):759. doi:10.3390/molecules30040759.
- 22. Mai H, Le TC, Chen D, Winkler DA, Caruso RA. Machine learning for electrocatalyst and photocatalyst design and discovery. Chem Rev. 2022;122(16):13478–515. doi:10.1021/acs.chemrev.2c00061.
- 23. Huang G, Guo Y, Chen Y, Nie Z. Application of machine learning in material synthesis and property prediction. Materials. 2023;16(17):5977. doi:10.3390/ma16175977.
- 24. Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman M, et al. A comprehensive overview of large language models. arXiv:2307.06435. 2023.
- 25. Chandrasekhar A, Farimani OB, Ajenifujah OT, Ock J, Farimani AB. NANOGPT: a query-driven large language model retrieval-augmented generation system for nanotechnology research. arXiv:2502.20541. 2025.
- 26. Polak MP, Morgan D. Extracting accurate materials data from research papers with conversational language models and prompt engineering. Nat Commun. 2024;15(1):1569. doi:10.1038/s41467-024-45914-8.
- 27. Miret S, Anoop Krishnan NM. Are LLMs ready for real-world materials discovery? arXiv:2402.05200. 2024.
- 28. Foppiano L, Lambard G, Amagasa T, Ishii M. Mining experimental data from materials science literature with large language models: an evaluation study. Sci Technol Adv Mater Meth. 2024;4(1):2356506. doi:10.1080/27660400. 2024.2356506.
- 29. Wu S, Ma X, Luo D, Li L, Shi X, Chang X, et al. Automated review generation method based on large language models. arXiv:2407.20906. 2024.
- Choi J, Im S, Choi J, Surendran S, Moon DJ, Kim JY, et al. Recent advances in 2D structured materials with defectexploiting design strategies for electrocatalysis of nitrate to ammonia. Energy Mater. 2024;4(2):400020. doi:10. 20517/energymater.2023.67.

- 31. Omeiza LA, Abdalla AM, Wei B, Dhanasekaran A, Subramanian Y, Afroze S, et al. Nanostructured electrocatalysts for advanced applications in fuel cells. Energies. 2023;16(4):1876. doi:10.3390/en16041876.
- 32. Ding R, Wang X, Tan A, Li J, Liu J. Unlocking new insights for electrocatalyst design: a unique data science workflow leveraging Internet-sourced big data. ACS Catal. 2023;13(20):13267–81. doi:10.1021/acscatal.3c01914.
- 33. Ali S, Ahmad A, Hussain I, Ahmad Shah SS, Sufyan Javed M, Ali S, et al. Experimental and theoretical aspects of MXenes-based energy storage and energy conversion devices. J Chem Env. 2023;2(2):54–81. doi:10.56946/jce. v2i2.214.
- 34. Pan Y, Shan X, Cai F, Gao H, Xu J, Zhou M. Accelerating the discovery of oxygen reduction electrocatalysts: high-throughput screening of element combinations in Pt-based high-entropy alloys. Angew Chem Int Ed. 2024;63(37):e202407116. doi:10.1002/anie.202407116.
- 35. Jijaba Kadam J, Ramchandra Jadhav M, Abasaheb Howal S, Martand Kharmate G, Uttam Pandit V. Machine learning-based optimization method for the oxygen evolution and reduction reaction of the high-entropy alloy catalysts. Int J Recent Innov Trends Comput Commun. 2023;11(9):2123–35. doi:10.17762/ijritcc.v11i9.9214.
- 36. Wei C, Shi Y, Mu W, Zhang H, Qin R, Yin Y, et al. Large language models assisted materials development: case of predictive analytics for oxygen evolution reaction catalysts of (oxy)hydroxides. ACS Sustain Chem Eng. 2025;13(14):5368–80. doi:10.1021/acssuschemeng.5c00798.
- 37. Zheng Z, Zhang O, Borgs C, Chayes JT, Yaghi OM. ChatGPT chemistry assistant for text mining and the prediction of MOF synthesis. J Am Chem Soc. 2023;145(32):18048–62. doi:10.1021/jacs.3c05819.
- 38. Akbashev AR. Electrocatalysis goes nuts. ACS Catal. 2022;12(8):4296–301. doi:10.1021/acscatal.2c00123.
- 39. Lin X, Du X, Wu S, Zhen S, Liu W, Pei C, et al. Machine learning-assisted dual-atom sites design with interpretable descriptors unifying electrocatalytic reactions. Nat Commun. 2024;15(1):8169. doi:10.1038/s41467-024-52519-8.
- 40. Gao Y, Wang L, Chen X, Du Y, Wang B. Revisiting electrocatalyst design by a knowledge graph of Cu-based catalysts for CO₂ reduction. ACS Catal. 2023;13(13):8525–34. doi:10.1021/acscatal.3c00759.
- 41. Huerta GV, Hisama K, Koyama M. A comprehensive computer-aided review of trends in heterogeneous catalysis and its theoretical modelling for engineering catalytic activity. J Chem Eng Jpn. 2024;57(1):2360928. doi:10.1080/00219592.2024.2360928.
- 42. Li X, Fu L, Chen F, Lv Y, Zhang R, Zhao S, et al. Cyclodextrin-based architectures for electrochemical sensing: from molecular recognition to functional hybrids. Anal Methods. 2025;17(21):4300–20. doi:10.1039/d5ay00612k.
- 43. Zhang C, Wang X, Wang Z. Large language model in electrocatalysis. Chin J Catal. 2024;59:7–14. doi:10.1016/S1872-2067(23)64612-1.
- 44. Wang L, Chen X, Du Y, Zhou Y, Gao Y, Cui W. CataLM: empowering catalyst design through large language models. Int J Mach Learn Cybern. 2025;16(5–6):3681–91. doi:10.1007/s13042-024-02473-0.
- 45. Park YJ, Jerng SE, Yoon S, Li J. 1.5 million materials narratives generated by chatbots. Sci Data. 2024;11(1):1060. doi:10.1038/s41597-024-03886-w.
- 46. Wang X, Huang L, Xu S, Lu K. How does a generative large language model perform on domain-specific information Extraction?—a comparison between GPT-4 and a rule-based method on band gap extraction. J Chem Inf Model. 2024;64(20):7895–904. doi:10.1021/acs.jcim.4c00882.
- 47. Eger S, Cao Y, D'Souza J, Geiger A, Greisinger C, Gross S, et al. Transforming science with large language models: a survey on AI-assisted scientific discovery, experimentation, content generation, and evaluation. arXiv:2502.05151. 2025.
- 48. Zimmermann Y, Bazgir A, Afzal Z, Agbere F, Ai Q, Alampara N, et al. Reflections from the 2024 large language model (LLM) hackathon for applications in materials science and chemistry. arXiv:2411.15221. 2024.
- 49. Yang Z, Liu W, Gao B, Xie T, Li Y, Ouyang W, et al. MOOSE-CHEM: large language models for rediscovering unseen chemistry scientific hypotheses. arXiv:2410.07076. 2024.
- 50. Song Z, Ju M, Ren C, Li Q, Li C, Zhou Q, et al. Leveraging large language models for universal scientific formula and theory discovery. arXiv:2503.06512. 2025.
- 51. Zheng Z, Florit F, Jin B, Wu H, Li SC, Nandiwale KY, et al. Integrating machine learning and large language models to advance exploration of electrochemical reactions. Angew Chem Int Ed. 2025;64(6):e202418074. doi:10.1002/anie. 202418074.

- 52. Su Y, Wang X, Ye Y, Xie Y, Xu Y, Jiang Y, et al. Automation and machine learning augmented by large language models in a catalysis study. Chem Sci. 2024;15(31):12200–33. doi:10.1039/d3sc07012c.
- 53. Duignan TT. The potential of neural network potentials. ACS Phys Chem Au. 2024;4(3):232-41. doi:10.1021/ acsphyschemau.4c00004.
- 54. Chen X, Gao Y, Wang L, Cui W, Huang J, Du Y, et al. Large language model enhanced corpus of CO₂ reduction electrocatalysts and synthesis procedures. Sci Data. 2024;11(1):347. doi:10.1038/s41597-024-03180-9.
- 55. Weston L, Tshitoyan V, Dagdelen J, Kononova O, Trewartha A, Persson KA, et al. Named entity recognition and normalization applied to large-scale information extraction from the materials science literature. J Chem Inf Model. 2019;59(9):3692–702. doi:10.1021/acs.jcim.9b00470.
- 56. Wang L, Gao Y, Chen X, Cui W, Zhou Y, Luo X, et al. A corpus of CO₂ electrocatalytic reduction process extracted from the scientific literature. Sci Data. 2023;10(1):175. doi:10.1038/s41597-023-02089-z.
- Kim E, Huang K, Saunders A, McCallum A, Ceder G, Olivetti E. Materials synthesis insights from scientific literature via text extraction and machine learning. Chem Mater. 2017;29(21):9436–44. doi:10.1021/acs.chemmater. 7b03500.
- 58. PEESEgroup/Awesome-materials-aware-large-language-models [Internet]. [cited 2025 Jun 11]. Available from: https://github.com/PEESEgroup/Awesome-Materials-Aware-Large-Language-Models.
- 59. Kim E, Jensen Z, van Grootel A, Huang K, Staib M, Mysore S, et al. Inorganic materials synthesis planning with literature-trained neural networks. J Chem Inf Model. 2020;60(3):1194–201. doi:10.1021/acs.jcim.9b00995.
- 60. Zuo Y, Chen C, Gong X, Zeng Y, Cai R, Ong SP. Accurate, interpretable predictions of materials properties within machine-learned interatomic potentials using explainable graph neural networks. npj Comput Mater. 2023;9(1):1–12. doi:10.1038/s41524-023-01095-4.
- 61. Ock J, Guntuboina C, Barati Farimani A. Catalyst energy prediction with CatBERTa: unveiling feature exploration strategies through large language models. ACS Catal. 2023;13(24):16032–44. doi:10.1021/acscatal.3c04956.
- 62. Shoghi N, Kolluru A, Kitchin JR, Ulissi ZW, Zitnick CL, Wood BM. From molecules to materials: pre-training large generalizable models for atomic property prediction. arXiv:2310.1680. 2023.
- 63. Pei Z, Yin J, Zhang J. Language models for materials discovery and sustainability: progress, challenges, and opportunities. Prog Mater Sci. 2025;145:14849. doi:10.1016/j.pmatsci.2025.101495.
- 64. Mok DH, Back S. Language model-based generative model for catalyst discovery Paper presented at: 2024 MRS Fall Meeting & Exhibit; 2024; Boston, MA, USA: Materials Research Society [Intermet]. [cited 2025 Jun 1]. Available from: https://www.mrs.org/meetings-events/annual-meetings/archive/meeting/presentations/ view/2024-fall-meeting/2024-fall-meeting-4148755.
- 65. Mok DH, Back S. Generative pretrained transformer for heterogeneous catalysts. J Am Chem Soc. 2024;146(49):33712-22. doi:10.1021/jacs.4c11504.
- Yu G, Mok DH, Jang HY, Jung HD, Siahrostami S, Back S. Leveraging Machine learning and active motifs-based catalyst design for discovery of oxygen reduction electrocatalysts for hydrogen peroxide production. J Catal. 2025;442(5):115906. doi:10.1016/j.jcat.2024.115906.
- 67. Zhang C, Lin Q, Zhu B, Yang H, Lian X, Deng H. SynAsk: unleashing the power of large language models in organic synthesis. Chem Sci. 2025;16(1):43–56. doi:10.1039/d4sc04757e.
- 68. Choi SE, Jang M, Yoon S, Yoo S, Ahn J, Kim M, et al. LLM-driven synthesis planning for quantum dot materials development. J Chem Inf Model. 2025;65(6):2748–58. doi:10.1021/acs.jcim.4c01529.
- 69. Pan Y, Xiao Q, Zhao F, Li Z, Liu J, Ullah S, et al. Chat-microreactor: a large-language-model-based assistant for designing continuous flow systems. Chem Eng Sci. 2025;311:121567. doi:10.26434/chemrxiv-2024-52g9l.
- Bandeira L, Ferreira H, de Almeida JM, Jardim de Paula A, Dalpian GM. CO₂ reduction beyond copperbased catalysts: a natural language processing review from the scientific literature. ACS Sustain Chem Eng. 2024;12(11):4411–22. doi:10.1021/acssuschemeng.3c06920.
- 71. Zitnick CL, Chanussot L, Das A, Goyal S, Heras-Domingo J, Ho C, et al. An introduction to electrocatalyst design using machine learning for renewable energy storage. arXiv:2010.09435. 2020.

- 72. Chen L, Tian Y, Hu X, Yao S, Lu Z, Chen S, et al. A universal machine learning framework for electrocatalyst innovation: a case study of discovering alloys for hydrogen evolution reaction. Adv Funct Mater. 2022;32(47):2208418. doi:10.1002/adfm.202208418.
- 73. Shan X, Pan Y, Cai F, Gao H, Xu J, Liu D, et al. Accelerating the discovery of efficient high-entropy alloy electrocatalysts: high-throughput experimentation and data-driven strategies. Nano Lett. 2024;24(37):11632–40. doi:10.1021/acs.nanolett.4c03208.
- 74. Phillips PJ, Phillips PJ, Hahn CA, Fontana PC, Yates AN, Greene K, et al. Four principles of explainable artificial intelligence. Natl Inst Stand Technol. 2021;9:1–23. doi:10.6028/nist.ir.8312.
- 75. Hysmith H, Foadian E, Padhy SP, Kalinin SV, Moore RG, Ovchinnikova OS, et al. The future of self-driving laboratories: from human in the loop interactive AI to gamification. Digit Discov. 2024;3(4):621–36. doi:10.1039/ d4dd00040d.