Materials Today Communications 40 (2024) 109801

Contents lists available at ScienceDirect

Materials Today Communications

journal homepage: www.elsevier.com/locate/mtcomm

Large language models in electronic laboratory notebooks: Transforming materials science research workflows

Mehrdad Jalali *, Yi Luo, Lachlan Caulfield, Eric Sauter, Alexei Nefedov, Christof Wöll *

Institute of Functional Interfaces (IFG), Karlsruhe Institute of Technology (KIT), Eggenstein-Leopoldshafen, Germany

ARTICLE INFO

Keywords:
Materials science research
Natural language processing (NLP)
Electronic laboratory notebooks (ELNs)
Large language models (LLMs)
Knowledge extraction
Scientific data management

ABSTRACT

In recent years, there has been a surge in research efforts dedicated to harnessing the capabilities of Large Language Models (LLMs) in various domains, particularly in material science. This paper delves into the transformative role of LLMs within Electronic Laboratory Notebooks (ELNs) for scientific research. ELNs represent a pivotal technological advancement, providing a digital platform for researchers to record and manage their experiments, data, and findings. This study explores the potential of LLMs to revolutionize fundamental aspects of science, including experimental methodologies, data analysis, and knowledge extraction within the ELN framework. We present a demonstrative showcase of LLM applications in ELN environments and, furthermore, we conduct a series of empirical evaluations to critically assess the practical impact of LLMs in enhancing research processes within the dynamic field of materials science. Our findings illustrate how LLMs can significantly elevate the quality and efficiency of research outcomes in ELNs, thereby advancing knowledge and innovation in materials science research and beyond.

1. Introduction

In recent years, Large Language Models (LLMs) have gained widespread attention as competent tools in artificial intelligence [1–3]. Initially known for their roles in applications such as chatbots, virtual assistants, and text generation, LLMs have rapidly expanded their influence across diverse fields. Scientific research promises to transform how experiments are conducted fundamentally, data is analyzed, and insights are generated. This paper embarks on a comprehensive exploration of LLMs in materials science, covering their fundamental concepts, practical applications, a comparative analysis of their effectiveness, and the development of specialized LLMs. We also delve into their integration into digital research environments like Electronic Laboratory Notebooks (ELNs) and provide insights into their practical use in streamlining research workflows. The paper culminates in an LLM-ELN demonstration, offering hands-on experience of these models' transformative potential. Finally, we conduct a comprehensive evaluation to assess their real-world impact and potential across scientific domains.

1.1. Large language model (LLM)

A Large Language Model is an artificial intelligence algorithm meticulously trained on vast amounts of text data to understand grammar, syntax, context, semantics, and word associations [1,4–6]. They excel in language-related tasks, such as translation, sentiment analysis, and question-answering. Prominent examples of LLMs include GPT-4 [7,8], BERT [9,10], and Transformer models [11], which have significantly advanced natural language processing capabilities.

1.1.1. LLMs in materials science

In materials science, where the quest for new materials with specific properties is constantly pursued, offer promising opportunities [12–17]. They can assist scientists in various critical tasks related to materials research, including data analysis, summarization of research papers, predicting material properties, and even generating novel material compositions. Researchers can expedite their work and make more informed decisions by leveraging the language model's profound understanding of materials-related content.

To provide an extensive overview of the existing research about large language models in this field, we investigated 200 papers in the Scopus repository as of May 2024. As depicted in Fig. 1, we conducted a word

* Corresponding authors.
E-mail addresses: mehrdad.jalali@kit.edu (M. Jalali), christof.woell@kit.edu (C. Wöll).

https://doi.org/10.1016/j.mtcomm.2024.109801
Received 14 May 2024; Received in revised form 24 June 2024; Accepted 7 July 2024
Available online 9 July 2024

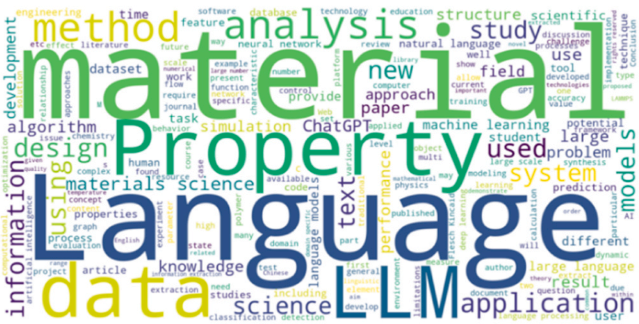

Fig. 1. Word cloud analysis of all the 200 papers.

cloud analysis of these papers. This analysis revealed that the predominant focus of current research lies in the domain of large language models in materials science. However, it is noteworthy that a substantial body of work also delves into bioinformatics, encompassing computational biology, protein conformation, gene ontology, drug discovery, and sequence alignment. Similarly, the domain of chemistry exhibits considerable attention, particularly in the study of metal-organic frameworks and material property prediction, among other facets. This observation is further emphasized by Fig. 2, which illustrates the distribution of submitted papers across diverse fields, thereby underscoring the demand for increased research and development efforts within these domains.

In our study on Large Language Models (LLMs) within the field of Materials Science, we harnessed the power of VOSviewer to construct a comprehensive network visualization. We extracted keywords from the titles and abstracts of 200 research papers, allowing us to delve deep into the intricate web of relationships between these keywords. To enhance the clarity of our visualization, we employed varying colors to highlight individual keywords along with their closely related counterparts within the same paper. This approach illuminated the central themes within each paper. It shed light on the interconnectedness of keywords across the entire corpus of research, providing a visually compelling representation of the multifaceted landscape of LLMs in Materials Science (Fig. 3).

1.2. Materials science language models

to better understand the potential of LLMs in materials science, it is essential to compare some of the leading models:

1. GPT-4 (Generative Pre-trained Transformer 4): Known for its versatility, GPT-4 can provide context-aware answers to complex questions related to materials science. Its ability to generate human-like text makes it valuable for summarizing research papers and explaining intricate scientific concepts [18–20].
2. BERT (Bidirectional Encoder Representations from Transformers): BERT’s bidirectional attention mechanism enables it to grasp context effectively, which is particularly useful for sentiment analysis in materials research. Understanding the sentiment behind research findings can help scientists gauge the impact and reception of their work [21].
3. Transformer Models: Transformer models, known for their self-attention mechanisms, excel in processing data sequences, which is common in materials science experiments. They can analyze the sequential data generated during experiments and identify patterns or anomalies [21–23].

1.3. Building private LLM for materials science

A domain-specific Large Language Model is a specialized variant of a large language model fine-tuned to excel in understanding and generating text related to a specific field or industry, such as healthcare [24–26], law [27], finance [28,29], or materials science [13,30], by learning the specialized terminology and context within that domain [31,32]. Developing and integrating a domain-specific LLM can be a transformative step in advancing research and innovation. This specialized LLM is tailored to the specific needs and complexities of the materials science domain, empowering researchers with advanced natural language understanding and capabilities. The process of creating and utilizing such a model involves several key steps. Data collection forms the foundation of a domain-specific LLM, as the quality and comprehensiveness of the training data directly impact its effectiveness. Researchers gather an extensive and diverse dataset comprising texts from reputable sources, including research papers, scientific journals, conference proceedings, and laboratory reports. Once assembled, this

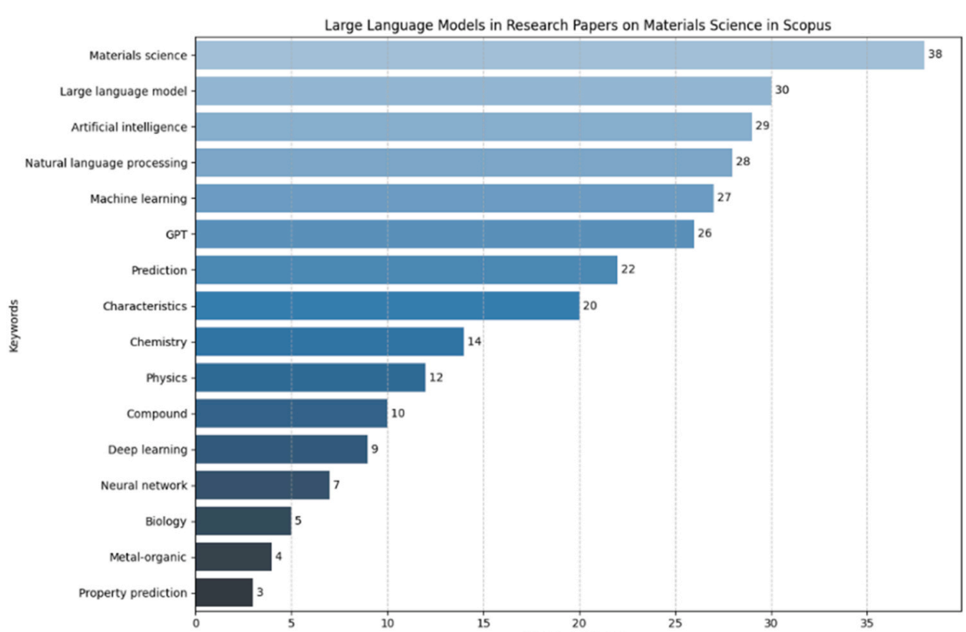

Fig. 2. shows the distribution of LLM papers submitted across various fields.

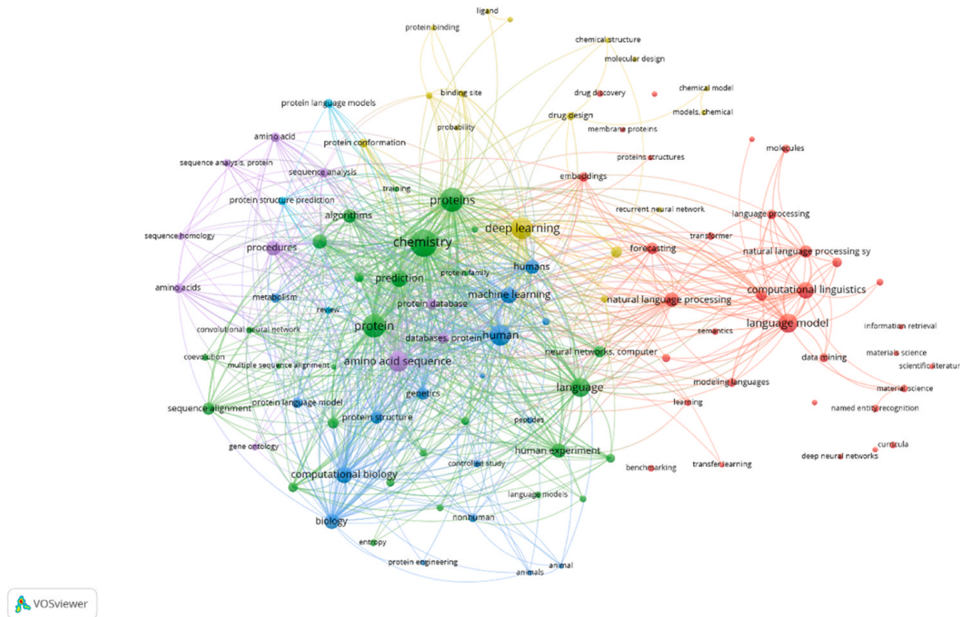

Fig. 3. Network Visualization of Keywords in LLM Research within Materials Science.

dataset undergoes meticulous preparation. This phase involves tokenization, dividing the text into meaningful units, such as words or subwords. Standardization and consistency across the dataset are crucial for practical training. The heart of domain-specific LLM development lies in the fine-tuning process. Researchers adapt a pre-existing language model to materials science by fine-tuning it on the prepared dataset. This step involves adjusting hyperparameters and extensive training to ensure that the LLM comprehends the nuances of materials science terminology and context. Ensuring data quality is paramount throughout this process. Researchers meticulously clean the dataset to eliminate irrelevant or noisy text. Any missing or inconsistent data is addressed to maintain a high standard of data quality. A clean dataset is essential for the model's accuracy and relevance. Incorporating a domain-specific LLM into materials science research signifies a significant leap in efficiency and effectiveness. This specialized model becomes a vital companion for researchers, aiding them in comprehending intricate concepts, predicting outcomes, and exploring new avenues in materials research. Its integration into the research workflow promises to streamline processes, expedite discoveries, and enhance the overall scientific journey in materials science. The flowchart of the process shows in Fig. 4.

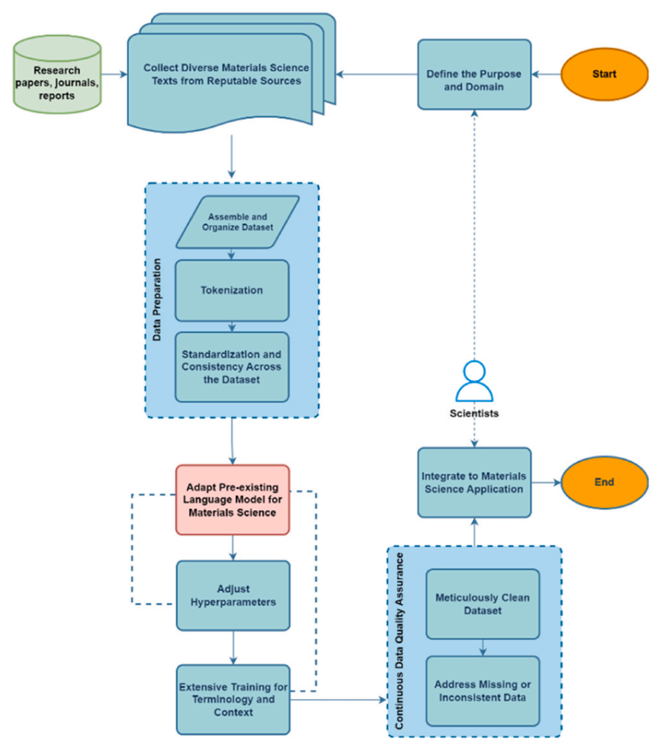

Fig. 4. Building Private LLM for Materials Science.

1.4. Applications of materials science language models

The applications of materials science language models are extensive and include:

1.4.1. Predicting material properties

Addressing the challenge of predicting material properties, the code in Fig. 5 utilizes a pre-trained Language Model (LLM) to classify the pore size of Metal-Organic Frameworks (MOFs) based on their descriptions. By defining categories and providing detailed descriptions for pore sizes, the code establishes a framework for automated classification. It then constructs a system prompt instructing the LLM on its task, which involves analyzing the MOF description and determining the appropriate pore size category. Leveraging the capabilities of the OpenAI GPT-4 Turbo model, the code initiates a chat-based interaction with the LLM, prompting it to generate a response predicting the pore size classification.

1.4.2. Discovery of the materials

The discovery of novel materials with desired properties is a central goal in materials science. LLMs can play a ↱pivotal role in this pursuit by assisting in identifying promising candidates. They can analyze vast datasets of ↱known materials and their properties to suggest potential materials that exhibit the desired characteristics. This ↱data-driven approach can significantly narrow down the search space for experimentalists, saving time and ↱resources [30,33,34]. For instance, to address the challenge of enhancing gas adsorption in Metal-Organic Frameworks (MOFs), materials scientists often seek innovative compositions. Large Language Models offer a solution by suggesting novel

```
import openai

# Set the API key
openai.api_key = 'your-api-key'

# Labels with descriptions for pore size classification
labels = {
    "non-porous": "0 nm - does not contain pores",
    "small pores": "less than 2 nm",
    "medium pores": "2 nm to 5 nm",
    "large pores": "greater than 5 nm"
}

# Example MOF description
mof_description = "Predict the pore size of a MOF with a framework incorporating linkers of 2 nm and a Zn4O (zinc tetraoxide) node."

# Create the system prompt including detailed labels for clarity in classification
label_text = ", ".join([f"[13] ({value})" for key, value in labels.items()])
system_prompt = f"You are a helpful assistant capable of classifying MOF pore sizes. Possible categories: {label_text}. Based on the description, classify the
pore size."

# Generate AI response
response = openai.ChatCompletion.create(
    model="gpt-4-turbo",
    messages=[
        {"role": "system", "content": system_prompt},
        {"role": "user", "content": mof_description}
    ]
)

# Print the AI's pore size classification
print(response.choices[0].message['content'])
```

LLM

```
Given the 2 nm size of the linkers and the presence of a typically robust and spatially accommodating Zn4O node, the MOF described is likely to
have **large pores**
```

Fig. 5. A Demonstrative python code for predicting material properties using large language models.

material combinations. Leveraging their extensive knowledge, LLMs propose compositions that may lead to breakthrough discoveries. In the provided Python code, we use the OpenAI API to utilize LLMs, specifically the GPT-4 Turbo model, for generating innovative MOF compositions aimed at enhancing gas adsorption (Fig. 6).

1.4.3. Intelligent properties extraction

One common challenge in materials science research is the extraction of relevant features and properties from scientific literature. Scientific texts often contain vast amounts of information presented in diverse formats and varying wordings. This diversity can make it challenging for researchers to efficiently extract specific details, such as the composition of Metal-Organic Frameworks (MOFs), their structural properties, or their performance characteristics. Additionally, nuances in language and terminology further complicate the extraction process, requiring a deep understanding of both the semantic and contextual aspects of the text. To address this challenge, intelligent extraction methods are necessary. These methods leverage advanced natural language processing (NLP) techniques to interpret the meaning and context of scientific texts. By employing sophisticated algorithms that can understand the nuances of language and recognize patterns in the data, researchers can automate the extraction process and efficiently retrieve the desired information from scientific literature.

For example, consider the code snippet provided in Fig. 7. It demonstrates an intelligent extraction approach using the GPT-4 Turbo model to extract specific properties of MOFs from a complex scientific text. The function extract_mof_properties takes as input a text description of a MOF and a list of queries regarding its properties. It then utilizes the GPT-4 Turbo model to interpret the queries in the context of the text and generate accurate responses.

1.4.4. Sentiment analysis in materials science

involves applying natural language processing techniques to evaluate the sentiment or emotional tone expressed in textual data related to materials, research findings, or reviews. This analysis can help researchers and engineers gauge the positivity or negativity of opinions, comments, or reviews about specific materials, material properties, or innovations. Sentiment analysis in materials science contributes to understanding researcher sentiment, assessing confidence levels in finding and identifying trends, monitoring public perception, supporting

```
# System prompt for generating innovative MOF compositions
system_prompt = """
As a materials science expert specializing in Metal-Organic Frameworks (MOFs), generate novel MOF compositions with high gas adsorption potential. Focus on
proposing unique combinations of metal nodes and organic linkers not extensively explored in existing literature.
"""

# Requesting AI response
response = openai.ChatCompletion.create(
    model="text-davinci-002",
    messages=[
        {"role": "system", "content": system_prompt},
        {"role": "user", "content": "I'm seeking innovative MOF compositions for enhanced gas adsorption. Can you suggest some ideas?"}
    ]
)

# Printing AI-generated innovative MOF composition ideas
print(response.choices[0].message['content'])
```

LLM

```
1. **Scandium-Trimesic Acid MOF**:
   - **Metal Node**: Scandium (Sc)
   - **Organic Linker**: Trimesic acid (benzene-1,3,5-tricarboxylic acid)
   - **Rationale**: Scandium-based MOFs are less studied despite scandium's potential for high coordination numbers and strong metal-linker
interactions. Pairing scandium with trimesic acid, a commonly used linker but underexplored with scandium, might yield a structure with high
porosity and exceptional stability.
2. **Thulium-Phthalocyanine MOF**:
   - **Metal Node**: Thulium (Tm)
 - **Organic Linker**: Phthalocyanine
   - **Rationale**: Phthalocyanines are macrocyclic compounds typically used in dyes and pigments but underutilized in MOFs. Incorporating these
with thulium, a lesser-used lanthanide in MOF applications, can potentially provide unique electronic environments conducive to gas interactions due
to the extensive π-network of the phthalocyanine.
3. **Yttrium-Naphthalenediimide MOF**:
   - **Metal Node**: Yttrium (Y)
   - **Organic Linker**: Naphthalenediimide (NDI)
   - **Rationale**: Yttrium is known for its high coordination chemistry, and when combined with NDI, which is known for its electron-accepting
properties, it might create MOFs suitable for capturing electron-rich gases like ammonia or sulfur-containing compounds.
```

Fig. 6. Demonstrative python code for generating ideas for generating innovative mof compositions for enhanced gas adsorption.

```python
def extract_mof_properties(text, queries):
    results = {}
    for query in queries:
        response = openai.ChatCompletion.create(
            model="gpt-4-turbo",
            messages=[
                {"role": "system", "content": "You are a scientific assistant knowledgeable about Metal-Organic Frameworks."},
                {"role": "user", "content": f"{query}\n\n{text}"}
            ]
        )
        results[query] = response.choices[0].message['content']
    return results

# Example complex text about MOFs
complex_text = """
In our latest study, we synthesized a novel MOF using copper (Cu) as the metal node and terephthalic acid as the linker. This framework, designated as Cu-MOF-1,
exhibits exceptional gas storage capabilities due to its unique structure. The pore limiting diameter (PLD) of Cu-MOF-1 has been measured to be approximately 3.2
Angstroms, which facilitates the adsorption and storage of small gas molecules such as hydrogen and methane. The structural integrity of Cu-MOF-1 under high
pressures and its reversible gas adsorption properties make it an ideal candidate for industrial applications.
"""

# Queries to extract specific information
queries = [
    "What is the PLD size mentioned?",
    "Identify the metal node used in the MOF.",
    "What type of linker is used in the MOF?"
]

# Extract MOF properties using GPT-4 Turbo
extracted_properties = extract_mof_properties(complex_text, queries)

# Print the results
for query, response in extracted_properties.items():
    print(f"{query}: {response}")
```

```
What is the PLD size mentioned?: The pore limiting diameter (PLD) mentioned for Cu-MOF-1 is approximately 3.2 Angstroms.
Identify the metal node used in the MOF.: The metal node used in the MOF described in your study, designated as Cu-MOF-1, is copper (Cu).
What type of linker is used in the MOF?: In the Metal-Organic Framework (MOF) you mentioned, known as Cu-MOF-1, the type of linker used is
terephthalic acid. Terephthalic acid, a common organic linker in the fabrication of MOFs.
```

Fig. 7. Demonstrative Python code for materials discovery.

decision-making, and assessing publication quality.

The provided Python code in Fig. 8 showcases the implementation of sentiment analysis within the realm of materials science utilizing a Large Language Model. Leveraging a pre-trained model, it discerns sentiment polarity, categorizing it as either "Positive" or "Negative," from supplied material review texts. This code illustrates the capacity of LLMs to aid researchers in autonomously gauging sentiment surrounding critical feedback on MOF Material Stability. The analysis output conveys a negative sentiment, indicating the expressed disappointment towards the stability of MOFs under operational conditions. Furthermore, it critiques the perceived lack of progress in fortifying their robustness, which it contends constrains their practical applications.

2. Integration of large language models into electronic lab notebooks (ELNs) for materials science

Electronic Lab Notebooks (ELNs) offer significant advantages when utilized with care in materials science. They play a crucial role in facilitating the systematic capture of data and information, ensuring consistency, accessibility, and usability for both current and future generations of scientists. ELNs can effectively address various challenges inherent in materials research. One prominent advantage is their alignment with open science principles, allowing for seamless data sharing within research organizations and compliance with the FAIR Data Principles [35,36]. These principles emphasize the importance of data being findable, accessible, interoperable, and reusable, which is essential in materials science for collaborative research and knowledge dissemination.

Furthermore, integrating ELNs with institutional data repositories presents exciting opportunities for enhancing research data management practices [37]. For instance, they enable researchers to directly deposit their data into institutional repositories through the ELN interface, simplifying the data archival process. ELNs also offer versatility by abstracting users from the intricacies of underlying notebook storage technologies. They empower scientists to implement decentralized record storage solutions, leveraging blockchain and peer-to-peer networking technologies [38]. This enhances data accountability and reduces reliance on a single repository for long-term data storage. By adopting this decentralized approach, data integrity verification, including laboratory notebook entries and scientific data, is distributed among consortium members, contributing to enhanced data security and trustworthiness.

Moreover, ELNs serve as the primary gateway to research data, opening opportunities for seamless integration with computational techniques. Researchers can explore the integration of ELNs with computational semantic technologies, which enable the automatic

```python
import openai

# Set the API key
openai.api_key = 'your-api-key'

# Example MOF comment
mof_comment = "The stability of MOFs under operational conditions remains a major disappointment. Despite years of research, there's little progress in
enhancing their robustness, which limits their practical applications."

# System prompt for sentiment analysis
system_prompt = "Analyze the sentiment of the following comment about MOF materials and classify it as positive, neutral, or negative."

# Perform sentiment analysis using GPT-4 Turbo
response = openai.ChatCompletion.create(
    model="gpt-4-turbo",
    messages=[
        {"role": "system", "content": system_prompt},
        {"role": "user", "content": mof_comment}
    ]
)

# Print the AI's sentiment analysis
print(response.choices[0].message['content'])
```

```
The sentiment of the comment about MOF materials is negative. The comment expresses disappointment regarding the stability of MOFs under
operational conditions and criticizes the lack of progress in enhancing their robustness, which it says limits their practical applications.
```

Fig. 8. Sentiment analysis of critical feedback on MOF material stability using GPT-4 turbo.

inference of human language's meaning. This capability has the potential to automatically derive research metadata, simplifying data retrieval and search efforts. Additionally, it can facilitate the creation of automated insights by connecting relevant data points, a valuable feature for scientists striving to uncover patterns and trends within their datasets. The meticulous implementation of ELNs within the field of materials science calls for a deep understanding of laboratory practices and a strategic approach to managing potential entry barriers. Successful integration of ELNs can significantly contribute to advancing knowledge within the science community. It is imperative to secure sustained institutional support, recognizing that the typical lifespan of ELN software packages is approximately seven years to ensure the ongoing benefits and continuity of research efforts in materials science [35].

Electronic Lab Notebooks (ELNs) have emerged as indispensable tools for enhancing data management, collaboration, and transparency in the fast-evolving landscape of scientific research. This state-of-the-art review examines prominent ELNs significantly contributing to various scientific disciplines, including Kadi4Mat, a research data infrastructure for materials science [39]. It empowers researchers to document experiments, manage data, and promote collaboration efficiently. ELOG [35] stands out for its robust data recording capabilities, ensuring meticulous documentation of experimental work. eLabFTW [40] offers a user-friendly interface that simplifies data organization and access. The groundbreaking introduction of the initial online electronic laboratory notebook, eCAT, was documented in reference [41]. This comprehensive analysis underscores the pivotal role of ELNs in modern scientific research and highlights their potential to shape the future of data-driven discovery.

Challenges in implementing ELNs in materials science are multifaceted. First, scientists grapple with integrating and standardizing a wide array of data types, ranging from complex chemical compositions to structural data and physical properties. This diversity makes maintaining a consistent data entry and nomenclature standard within the ELN a formidable challenge. Additionally, integrating specialized instruments, vital for many experiments, demands technical expertise to ensure seamless data capture and compatibility. Moreover, the need to safeguard sensitive research data and proprietary information heightens the importance of robust data security measures, adding an extra layer of complexity. Furthermore, a persistent challenge lies in addressing scientists' perceptions that the information kept in ELNs is often underutilized and overlooked, leading to the belief that ELNs can be burdensome and uninspiring.

The integration of Large Language Models into Electronic Laboratory Notebooks revolutionizes materials science research, offering unparalleled data management, analysis, and collaboration capabilities, significantly streamlining research workflows and fostering cross-disciplinary communication (see Table 1). LLMs address the challenge of efficient search and retrieval of information within ELNs by enabling quick and accurate searches for specific experiments and materials properties, enhancing accessibility and informed decision-making in materials development. Additionally, LLMs facilitate contextual data analysis, deciphering complex ELN entries and transforming them into understandable insights tailored to the materials domain. Moreover, LLMs accelerate the documentation process in materials science research by offering autocompletion and contextual assistance with materials-specific terminology, allowing researchers to focus more on experimentation and analysis. LLMs demonstrate their predictive analytics capabilities by predicting critical outcomes based on historical ELN data, aiding in informed decision-making for future experiments and material design. Furthermore, LLMs bridge the gap between ELN data and science literature, helping researchers identify related work, stay updated with advancements, and discover new areas for exploration and innovation. In terms of collaboration, LLMs enable dispersed research teams to collectively analyze ELN data, fostering knowledge sharing and ensuring a shared understanding of results crucial for advancing materials research. Data security and ethics are ensured with LLMs, which can automatically redact or secure sensitive information related to proprietary materials formulations and intellectual property, safeguarding the integrity of materials research. To maximize the benefits of LLMs in materials science, fine-tuning and training on domain-specific ELN data are essential, enhancing their performance within the materials domain. Despite their promise, challenges such as mitigating biases in LLM predictions require active research to ensure the reliability and ethical use of these advanced tools. In conclusion, the integration of LLMs into ELNs revolutionizes materials science research, enhancing efficiency, knowledge sharing, and prediction capabilities, renewing enthusiasm for exploration and pushing the boundaries of the field. Essential prerequisites for successful adoption include hardware, software, data

Table 1
Integration of large language models into electronic laboratory notebooks for materials science research.

Aspect	Description
Revolutionizing Research	**Data Management & Analysis** - Streamlined research workflows and cross-disciplinary communication through improved data management and analysis capabilities - Fostering collaborative research and enhancing data retrieval
Search & Retrieval	**Efficient Searches** - Quick and accurate searches for specific experiments and material properties within ELNs - Improved decision-making by enhancing data accessibility
Contextual Analysis	**Enhanced Understanding** - Deciphering complex ELN entries into understandable insights specific to materials science - Tailored analysis to suit the materials domain
Documentation Assistance	**Autocomplete & Terminology Support** - Contextual assistance with materials-specific terminology, improving documentation quality - Researchers can focus more on experimentation and analysis
Predictive Analytics	**Outcome Predictions** - Predicting critical outcomes based on historical ELN data - Aiding in informed decision-making for future experiments and material design
Bridging ELN Data & Literature	**Related Work Identification** - Connecting ELN data with related materials science literature - Discovering new areas for exploration and innovation
Collaboration	**Knowledge Sharing** - Collective ELN data analysis among dispersed research teams - Ensuring a shared understanding of results, crucial for advancing materials research
Security & Ethics	**Data Security** - Automatic redaction and securing sensitive proprietary formulations and intellectual property - Safeguarding the integrity of materials research
Customization & Fine-Tuning	**Domain-Specific Performance** - Training and fine-tuning on domain-specific ELN data - Enhancing performance within the materials domain
Challenges	**Bias Mitigation** - Addressing biases in LLM predictions - Ensuring reliability and ethical use
Adoption Prerequisites	**Hardware & Software** - Robust hardware, software, data storage, computational resources, and time - Facilitating the development and utilization of LLMs in materials science research
Conclusion	**Revolutionizing Materials Research** - Enhancing efficiency, knowledge sharing, and predictive capabilities - Renewing exploration enthusiasm and pushing the field's boundaries

storage, computational resources, and time, each playing a distinct role in facilitating the development and utilization of these powerful language models within this research field.

3. Approach for incorporating LLMs in the eLabFTW ELN

Integrating a Language Model into an Electronic Laboratory Notebook such as eLabFTW [42] offers a transformative approach to research workflows, as illustrated in the diagram. The workflow highlights the seamless interaction between internal data storage, text analysis, and ELN content via a central LLM system. This allows researchers to input natural language queries and receive detailed responses directly. For instance, the HKUST-1 experiment is described with recommendations on how long the sample should be left in the oven at 425°C to remove all water and carbonyls effectively (see Fig. 9).

The four critical phases involved in this integration are examined in detail.

3.1. Pre-integration planning

Integrating LLMs into an ELN begins with thorough pre-integration planning to ensure alignment with the laboratory's specific needs and objectives. This involves a comprehensive needs assessment to identify areas where LLMs can add value, such as automating data interpretation or simplifying documentation. The selection of an appropriate LLM model that aligns with laboratory requirements is crucial, alongside securing authorized API access as the gateway for data exchange between eLabFTW and the LLM.

3.2. Data handling and interaction

Efficient data handling and interaction form the core of LLM integration within eLabFTW. Experimental data is gathered, structured, and preprocessed for integration, while user-friendly interfaces empower researchers to interact seamlessly with the LLM. These interfaces allow them to input data or queries and receive responses in natural language. The LLM analyzes the data, generates insights, and automates documentation, leveraging natural language processing to aid researchers in making informed decisions.

3.3. Quality assurance and compliance

Quality assurance and compliance ensure integration reliability, security, and compliance. Rigorous testing validates performance, while robust data security measures safeguard sensitive experimental data. Ethical considerations include responsible AI usage and compliance with relevant regulations. Scalability measures accommodate future data growth, and regular maintenance keeps the integration up-to-date and adaptable.

3.4. Deployment and continuous improvement

The deployment phase emphasizes user training and documentation to facilitate widespread access to the LLM features within eLabFTW. Continuous monitoring of metrics like response times and accuracy ensures the integration meets expectations. An iterative development approach driven by user feedback guarantees adaptability. By embracing this comprehensive workflow, laboratories can harness the transformative potential of LLMs, elevating research processes and the overall quality of scientific work.

This study used the GPT-4 model as a pre-trained language model to enhance our research in ELNs. Although we also explored fine-tuning using GPT-3.5, the results were suboptimal due to the limited availability of ELN data, which is crucial for effective fine-tuning. Insufficient data can lead to poorer performance than solely using a robust pre-trained model like GPT-4. For future work, we plan to fine-tune using more extensive datasets and explore models such as LLaMA, which may provide improved performance if more comprehensive ELN data becomes available. This will allow us to better leverage the capabilities of these models for our specific domain needs.

4. Application of generated response by LLM in ELN

We categorized the generated responses and applications of using LLMs in ELNs into three categories: "Experiment Information," "Experimental Parameters," and "Documentation and Reports," to streamline research workflows and maintain high documentation standards.

4.1. Experiment information

This category emphasizes clarity, accountability, and effective communication in laboratory research. By defining sub-categories like

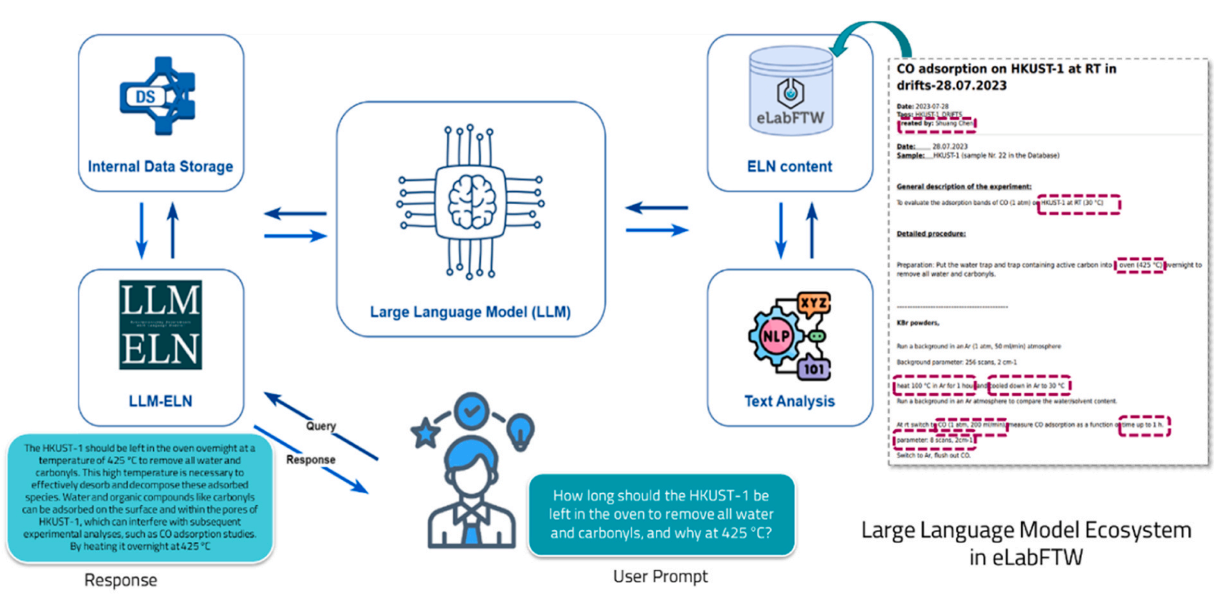

Fig. 9. Illustration of Large Language Model Ecosystem in eLabFTW.

Experimenter (e.g., Shuang Chen and THEO team for the HKUST-1 experiment), Objective (evaluating CO adsorption bands on HKUST-1 at room temperature), and Summary (concise experiment overview), researchers ensure clear attribution and research direction. The Usability sub-category provides practical applications, such as understanding HKUST-1's role in environmental filtration. Experiment Re-Planning offers suggestions to improve future experiments, like verifying initial conditions and statistical analysis, and enhancing replicability and reliability.

4.2. Experimental parameters

Comprehensive documentation is essential for replicability. Defining the Procedure Stage (e.g., oven cleaning, sample heating, and cooling), the Sample (HKUST-1, a metal-organic framework composed of copper ions and BTC ligands), the Equipment (oven, DRIFTS cell, LN MCT Detector, Microscopes), and Safety (using water and active carbon traps, maintaining an Argon atmosphere) provide a clear experimental roadmap. This ensures consistent results, minimizes errors, and maintains safety standards, as shown in the detailed HKUST-1 experiment list. Integrating

4.3. Documentation and reports

Maintaining quality and consistency in documentation and reporting is crucial for research credibility. Report Writing emphasizes generating high-quality reports reflecting the experimental process and findings (e.g., successful CO adsorption on HKUST-1). Documentation Standards ensure clarity, organization, and compliance, enabling efficient peer review and regulatory compliance. For instance, detailed safety protocols like using Ar for leak checks and flushing post-CO adsorption illustrate adherence to rigorous standards.

By thoughtfully categorizing generated answers into these segments, laboratories can streamline their data management, improve researcher communication, and maintain high standards of documentation integrity. Integrating LLMs into ELNs enhances research processes and elevates scientific endeavors' quality and transparency.

The program we've developed is Chabot for researchers and scientists, streamlining laboratory research by providing tailored AI-driven responses to user queries within Electronic Laboratory Notebook documents (Fig. 10). Seamlessly integrating Large Language Models into the ELN environment offers efficient access to essential information without manual document searches.

The user-friendly interface includes a text box for inputting questions, and by clicking the submit button, users promptly receive accurate responses. A noteworthy feature is the word cloud, a dynamic visual summary of frequently occurring keywords in ELN documents (Fig. 11). This visual aid complements detailed textual responses, simplifying information retrieval and data navigation, benefiting research productivity and decision-making. The program, exemplified by the word cloud, enhances usability and efficiency, empowering researchers to harness LLM capabilities effectively.

In this demonstration, we showcase the integration of Large Language Models into Electronic Laboratory Notebooks to streamline research processes and provide instant access to critical information. By addressing specific user inquiries, we illustrate how ELNs equipped with LLM capabilities can enhance data management and knowledge retrieval for researchers. In this case study, we applied Large Language Model to streamline the workflow for an experiment investigating carbon monoxide adsorption on the metal-organic framework HKUST-1. Here, we present specific examples of LLM-generated answers, highlighting their relevance and benefits.

4.4. Primary objective of the experiment

When asked about the primary objective of the experiment, the LLM responded concisely. This direct answer facilitates efficient communication of experimental goals, making it easier to summarize findings

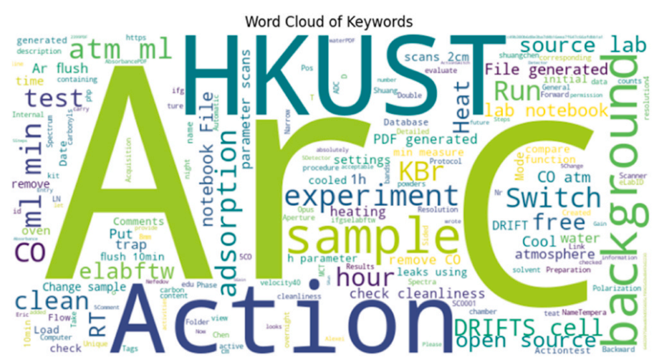

Fig. 11. Presentation of the most influential tokens in LLM-ELN Interaction. Word clouds highlight tokens with a significant impact on responses generated by our program. Tokens' font sizes represent their influence, and color differentiation aids in distinguishing closely related terms.

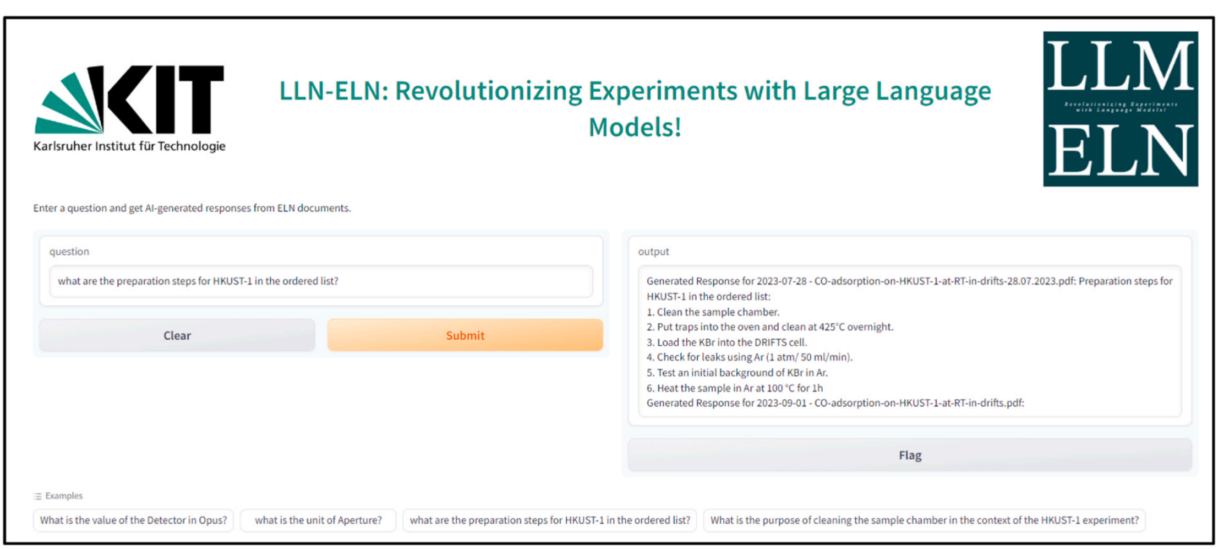

Fig. 10. Snapshot of the LLM-ELN application.

and align research directions.

The primary objective of the experiment was to evaluate the adsorption bands of CO on HKUST-1 at room temperature.

4.5. Sample heating procedure

Another prompt addressed the rationale behind heating samples in an argon (Ar) atmosphere; this comprehensive answer clarifies the preparation procedures, helping researchers maintain consistency and ensure reproducibility.

The reason for heating the samples in an Ar atmosphere and then cooling them down is primarily to clean the sample by removing any residual water solvent traces or adsorbed contaminants. This ensures that the sample surface is clean and suitable for accurate subsequent analysis of CO adsorption without interference from other substances.

We organize procedural details into Table 2 for clarity.

4.6. Usability of the experiment

The LLM elaborated on the usability of the CO adsorption experiment; these insights demonstrate how LLMs can synthesize potential applications and benefits, contributing to a comprehensive understanding of the experiment's relevance.

Material Science and Chemistry: Understanding HKUST-1's adsorption of CO is crucial for selective gas filtration, especially in CO-rich environments.
Environmental Applications: HKUST-1's CO adsorption at room temperature could aid air purification in enclosed spaces like parking lots or industrial areas.
Industrial Safety: Industries producing CO could use HKUST-1 for active adsorption, reducing CO poisoning risk.
Research and Development: This research advances metal-organic framework (MOF) applications in gas storage and catalysis, improving HKUST-1's efficiency and selectivity.
Academic and Scientific Insight: Results offer a baseline for comparing CO adsorption capacities among different materials, informing academic discussions on gas-solid interactions.

We have tabulated the usability insights generated by the LLM in Table 3.

4.7. Suggestions for experiment optimization

When prompted for suggestions to optimize the experiment, the LLM generated valuable recommendations, these suggestions, tailored to the experimental setup, demonstrate how LLMs can guide researchers in enhancing their methodologies.

4.7.1. Increasing replicability

Provide detailed procedural steps including equipment brand and model to ensure reproducibility.

Table 2
Heating procedure for MOF preparation.

Step	Description
Heating Atmosphere	Argon (Ar)
Heating Temperature	100°C
Heating Time	1 Hour
Purpose	Remove residual water solvent traces or contaminants
Final Temperature	Room Temperature (30°C)

Table 3
Application areas of MOFs in CO adsorption.

Application Area	Details
Material Science	Selective gas filtration in CO-rich environments
Environmental Applications	Air purification in enclosed spaces like parking lots
Industrial Safety	Reduce CO poisoning risk in CO-producing industries
Research and Development	Advances MOF applications in gas storage and catalysis
Academic and Scientific	Establishes baseline for CO adsorption capacity comparison

4.7.2. Verification of initial condition

Confirm equipment cleanliness with control samples, including blank runs without any sample, to establish baseline conditions before the experiment.

Sample Preparation Consistency:

Specify sample preparation details such as the weight or volume of HKUST-1 used to ensure consistency across trials and minimize variability.

4.7.3. Statistical analysis

Perform multiple replicates of the experiment to enable robust statistical analysis, affirming the reliability of findings by understanding data variation.

4.7.4. Temperature control

Monitor and record temperature at various stages of the experiment, not just after heating or cooling, to mitigate the impact of inconsistent temperatures on results.

Table 4 presents optimization suggestions in a table for quick reference.

4.8. Equipment details and settings

The LLM also provided detailed information about the equipment used; providing this equipment information helps researchers accurately replicate the experimental conditions.

Table 4
Recommendations for Improving Experiment Quality.

Aspect	Suggestion
Replicability	Provide detailed procedural steps including equipment brand and model
Initial Condition	Confirm equipment cleanliness with control samples
Sample Preparation	Specify weight/volume of HKUST-1 used
Statistical Analysis	Perform multiple replicates for robust statistical analysis
Temperature Control	Monitor and record temperature at various stages

4.8.1. Oven

For cleaning water traps and trap with active carbon at 425 °C.

4.8.2. Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS) Cell

Used for measuring CO adsorption on HKUST-1.

4.8.3. Liquid nitrogen mercury-cadmium-telluride (LN-MCT) detector

Narrow internal position 2 for detecting the spectra.

4.8.4. Computer setup

Used for controlling experiments and data storage, with specifics located at D:\data\shuangchen\20230728-HKUST-1.

Table 5 summarizes the equipment settings in a tabular format.

4.9. Safety guidelines

The LLM also identified critical safety rules mentioned in the experiment.

Use of water and active carbon traps to remove all water and carbonyls.
Operation in Argon atmosphere to maintain an inert environment and ensure no exposure to CO during critical stages.
Leak checks using Argon to ensure system integrity and avoid gas leaks.
Flushing system with Argon post-exposure to completely remove CO.
Cooling protocol to manage temperature and reduce thermal hazards.
Cleaning of DRIFTS cell to prevent contamination and maintain equipment integrity.

4.10. LLM-generated predictions and property insights

The LLM synthesizes data to provide an approximate adsorption capacity, reducing the need for extensive literature searches. This aids in quick comparisons and enhances data interpretation. In the following, we showcase three examples to demonstrate how it works.

4.10.1. What is the CO adsorption capacity of HKUST-1 based on ELN data?

(continued on next column)

(continued)

HKUST-1 exhibits high adsorption capacity due to its porosity and surface area. Based on ELN data, the adsorption capacity of CO on HKUST-1 is estimated at 4.5 mmol/g. This aligns with previous research findings for MOF-199 structures.

4.10.2. What does the CO adsorption isotherm look like for HKUST-1 at room temperature?

The CO adsorption isotherm on HKUST-1 at RT follows a Type I isotherm. This suggests a strong affinity between HKUST-1 and CO, with saturation occurring rapidly at low pressures. The estimated Langmuir surface area is around 1600 m^2/g.

4.10.3. How does HKUST-1 behave during the desorption process of CO?

During the desorption process, HKUST-1 releases adsorbed CO efficiently when flushed with Ar. The desorption process occurs rapidly within 10 minutes, indicating weak physisorption. The observed decrease in absorbance at 1400–1700 cm^{-1} supports this behavior.

Table 6 Present the predictions and insights using a table for easy comparison.

4.11. LLM-Generated Sentiment Analysis Examples

By categorizing the comments into positive and neutral sentiments, the LLM provides insight into the experimenter's confidence in the experimental setup and results. Positive comments suggest satisfaction with the preparation steps and adherence to protocols, while neutral comments indicate routine procedural steps.

4.11.1. What is the overall sentiment of the comments made by the experimenter during the HKUST-1 CO adsorption experiment?

The overall sentiment of the experimenter's comments is positive. Below are some notable comments with their associated sentiments; Positive: Now it looks absolutely acceptable. We, with Eric, checked the Entry; you have permission to carry out all activities corresponding to this experiment.

The integration of Large Language Models into Electronic Laboratory Notebooks showcases their potential to transform the way researchers access and interpret experimental data. In this case study, the LLM effectively provided answers, clarified procedures, suggested improvements, and offered safety guidelines, highlighting the value of LMs in streamlining research workflows. By enhancing data management, knowledge retrieval, and procedural optimization, LLMs in ELNs empower researchers to conduct more efficient, accurate, and reproducible experiments. This seamless blend of technology and research methodology promises to redefine laboratory practices, driving innovation and fostering deeper insights into scientific inquiry.

Table 5
Equipment and usage details.

Equipment	Details
Oven	Cleaning water traps and trap with active carbon at 425°C
DRIFTS Cell	Measuring CO adsorption on HKUST-1
LN-MCT Detector	Narrow internal position 2, for detecting spectra
Computer Setup	D:\data\shuangchen\20230728-HKUST-1

Table 6
Predicted answers by LLM.

Question	LLM-Generated Answer
CO Adsorption Capacity	4.5 mmol/g, aligning with previous research on MOF-199 structures
CO Adsorption Isotherm	Type I isotherm at room temperature (RT)
Desorption Behavior of HKUST-1	Rapid CO desorption within 10 minutes, weak physisorption

5. Evaluation

In this section, we analyze the effectiveness of proposed models using the Cosine Similarity Score and Semantic Similarity Score [43], alongside assessing the time complexity of LLM-ELN during inference phase. Our comprehensive assessment includes comparing LLM outputs with direct responses from ChatGPT, utilizing both expert human and automated metrics to determine performance accuracy and computational efficiency in real-time applications, such as processing PDF files and ELN documents. This evaluation is crucial for understanding the specific research needs within the eLabFTW environment and optimizing LLMs for enhanced research support.

5.1. Cosine similarity score

Cosine similarity measures how closely the generated responses (Generated) align in meaning with the reference responses (Reference). It is calculated by finding the cosine of the angle between the vector representations of the two texts, resulting in a score ranging from −1–1. In practice, the score is typically normalized to range from 0 (no similarity) to 1 (perfect similarity).

$$Cosine\ Similarity(Generated, Reference) = \frac{Generated.Reference}{\|Generated\|.\|Reference\|}$$

Cosine similarity directly compares two texts without relying on advanced word embeddings or contextual models. Instead, it relies on simple frequency-based vectors or other basic text representations.

5.2. Semantic similarity score

Semantic similarity measures how closely two texts align in meaning. It is a useful metric for evaluating the conceptual relationship between texts, as it considers the context and semantics rather than focusing solely on exact word matches. By analyzing the meaning and concepts behind words, semantic similarity provides a score ranging from 0 (no similarity) to 1 (identical meaning).

This metric can be calculated using word embeddings or contextual language models, which represent words or phrases in a high-dimensional vector space based on their semantic relationships. The similarity score, often represented as Sim(Generated, Reference), indicates how well the generated responses (Generated) match the reference responses (Reference) in terms of meaning and content. Word vectors or contextual embeddings are used to represent text meaningfully (Table 7).

$$Semantic\ Similarity(Generated, Reference) = \frac{Emb\_Generated.Embd\_Reference}{\|Emb\_Generated\|.\|Emb\_Reference\|}$$

The comparative analysis of responses from LLM-ELN, Expert, and GPT reveals that LLM-ELN, which directly utilizes experimental data, consistently provides more accurate and detailed responses than both the Expert and GPT. Across all questions, LLM-ELN responses show high semantic similarity scores with Expert responses, indicating strong conceptual alignment, while low cosine similarity scores highlight terminology differences.

For instance, in the question "How long does it take for CO to be absorbed by HKUST-1 at room temperature?", the LLM-ELN response provided a comprehensive, data-driven answer closely aligned with the Expert's concise response. In contrast, GPT's more general answer, not grounded in specific experimental data, showed a lower similarity with the Expert.

Similarly, for the question "How is HKUST-1 cleaned before testing?", LLM-ELN presented a detailed seven-step cleaning process that closely aligned with the Expert response, while GPT focused more on techniques affecting CO absorption, resulting in lower similarity.

Overall, integrating large language models with Electronic Laboratory Notebooks enhances the accuracy and depth of scientific insights. Fig. 12 illustrates this comparative analysis. Thus, it was shown that LLM-ELN consistently provides more accurate and detailed answers than the Expert and GPT. In the four example queries, LLM-ELN responses have high semantic similarity with Expert responses but low cosine similarity, highlighting terminology differences.

The time complexity of LLMs during inference is influenced by the specific architecture of the model. For example, transformer-based models such as GPT-4 exhibit a time complexity of O(nˆ2) per layer, where n represents the sequence length. This quadratic complexity is a

Table 7
Comparison of responses from LLM-ELN, Expert, and GPT in terms of cosine and semantic similarities, demonstrating the accuracy and detail provided by LLM-ELN responses.

No	Question	LLM-ELN	Expert	GPT	Cosine Similarity			Semantic Similarity		
					LLM-ELN - Expert	LLM-ELN - GPT	Expert - GPT	LLM-ELN - Expert	LLM-ELN - GPT	Expert - GPT
Q1	How long does it take for CO to be absorbed by HKUST-1 at RT?	The experiment as detailed in the logs indicates that CO was allowed to be absorbed by HKUST-1 for up to 1 hour at room temperature.	Based on the ELN data it takes about 1 hour.	The absorption rate of CO by HKUST-1 can vary based on several factors such as concentration, surface area, and conditions.	0.1575	0.4597	0.1190	0.6978	0.8583	0.5611
Q2	How is HKUST-1 cleaned before testing?	Describes a detailed 7-step cleaning process, including equipment setup and experimental conditions.	Washing with EtOH, heating for X amount of time at 100°C in Ar stream or vacuum.	Discusses factors affecting CO absorption and various experimental techniques.	0.1575	0.1898	0.0660	0.7768	0.8573	0.8121
Q3	Can CO adsorption results on HKUST-1 be reproduced?	Detailed explanation of reproducibility factors, including equipment settings, environmental conditions, and sample consistency.	Yes, if the conditions are being kept the same.	Reproducing CO adsorption results on HKUST-1 under identical conditions is theoretically possible but can be challenging.	0.1575	0.4597	0.1190	0.6978	0.8583	0.5611
Q4	What are the main findings from the CO adsorption experiment?	Describes the methodology and preparation steps but lacks specific findings.	Findings depend on the type of experiment performed.	Provides a comprehensive overview of CO adsorption, including capacity, kinetics, and thermodynamics.	0.2953	0.4011	0.2952	0.8226	0.9384	0.9015

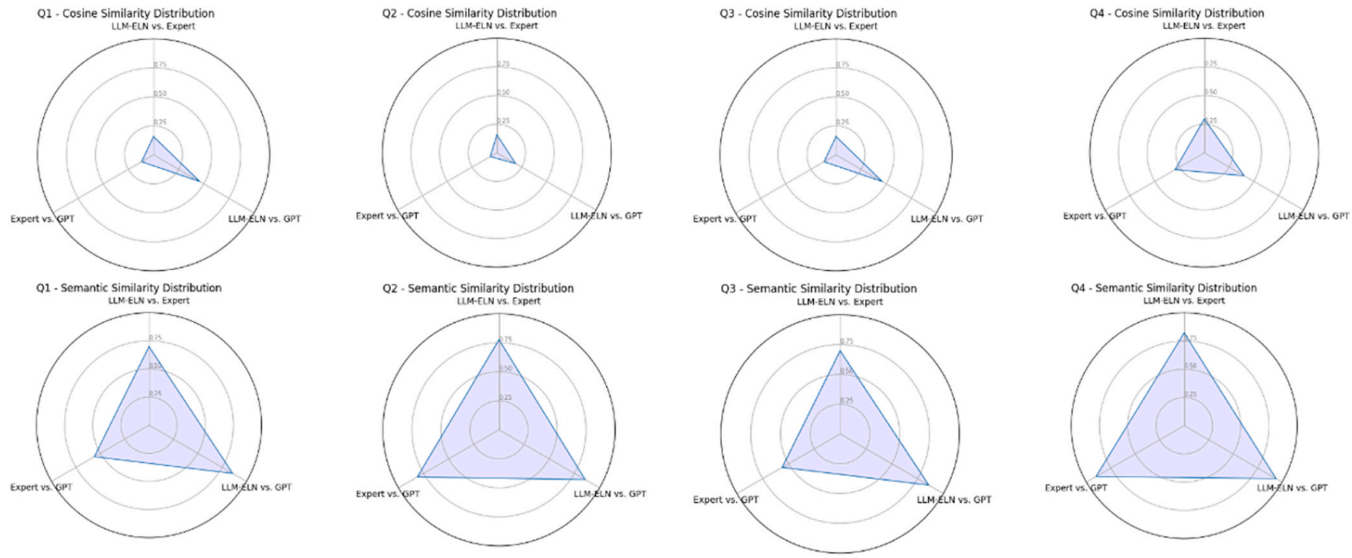

Fig. 12. The comparative analysis of responses from LLM-ELN, Expert, and GPT.

result of the self-attention mechanisms integral to transformers. In our implementation, where LLM-ELN is used for processing PDF files in real-time, the time complexity is significantly affected by both the length and the number of ELN documents being processed. This aspect is critical in determining the feasibility and efficiency of using LLM-ELN in real-world applications where performance and speed are crucial.

6. Conclusion and future direction

This study explored the integration of Large Language Models (LLMs) with Electronic Laboratory Notebooks (ELNs), highlighting their transformative potential in scientific research. Our evaluation demonstrated that the LLM-ELN system significantly enhances data retrieval, documentation, and interpretation, making the research process more efficient and enabling data-driven decision-making. Using metrics such as Cosine Similarity and Semantic Similarity, we assessed the model's performance, revealing that LLM-ELN responses closely aligned with expert responses in terms of semantic content despite some differences in terminology. Furthermore, we examined the time complexity of LLMs during inference, which is influenced by the length and number of ELN documents processed.

Looking ahead, several vital areas warrant further exploration. First, enhancing the dataset size and diversity will be critical for improving the fine-tuning capabilities of models like GPT-4. Expanding the ELN data repository to include more comprehensive and varied experimental data could enhance the model's accuracy and reliability. Developing user-friendly interfaces and tools that facilitate seamless interaction between LLMs and ELNs will also be essential for broader adoption in the scientific community.

We plan to enhance the integration of outputs from LLM-ELN into repositories like the Novel Materials Discovery (NOMAD) repository [44,45]. This initiative will promote open science and provide global access to essential research data. The process involves extracting and organizing experimental data from ELNs using the LLM-ELN system and then formatting it to meet NOMAD's standards for seamless integration. Additionally, we aim to refine this integration by incorporating the MSLE ontology [46], which effectively defines the semantics of materials science laboratory equipment. This strategic enhancement will standardize and enrich ELN metadata, aligning it with MSLE's framework to improve data sharing, interoperability, and analysis across platforms like NOMAD. This approach leverages the strengths of both MSLE and LLM-ELN to optimize scientific data management.

Integrating Large Language Models (LLMs) with Electronic Laboratory Notebooks (ELNs) marks a significant advancement in scientific research. By refining these technologies and expanding their applications, we can significantly enhance the efficiency, transparency, and impact of scientific discovery, driving breakthroughs across various fields. Collaborative efforts among researchers, AI specialists, and domain experts are essential to unlocking the full potential of LLMs, revolutionizing laboratory research practices, and fostering innovation.

CRediT authorship contribution statement

Lachlan Caulfield: Formal analysis, Data curation. **Eric Sauter:** Writing – review & editing, Formal analysis. **Yi Luo:** Formal analysis, Data curation. **Mehrdad Jalali:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Alexei Nefedov:** Writing – review & editing, Resources, Formal analysis. **Christof Wöll:** Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of Competing Interest

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

Data availability

https://github.com/MehrdadJalali-KIT/LLM-ELN

Acknowledgments

This project was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) in the framework of the project FAIRmat (Project ID: 460197019). The authors thank Dr. Shuang Chen for permission to use her eLabFTW entry.

Code and data availability

All datasets and custom code supporting this study's findings are openly available in the GitHub repository at github.com/MehrdadJalali-KIT/LLM-ELN. This repository contains the complete datasets used, the source code for data analysis and visualization, and detailed documentation to ensure reproducibility and further research. For additional information or specific requests, please contact the corresponding author.

References

[1] Shen, Y., Heacock, L., Elias, J., Hentel, K.D., Reig, B., Shih, G., and Moy, L.: 'ChatGPT and other large language models are double-edged swords', in Editor (Ed.)^(Eds.): 'Book ChatGPT and other large language models are double-edged swords' (Radiological Society of North America, 2023, edn.), pp. e230163.

[2] K. Singhal, S. Azizi, T. Tu, S.S. Mahdavi, J. Wei, H.W. Chung, N. Scales, A. Tanwani, H. Cole-Lewis, S. Pfohl, Large language models encode clinical knowledge, Nature (2023) 1–9.

[3] T. Webb, K.J. Holyoak, H. Lu, Emergent analogical reasoning in large language models, Nat. Hum. Behav. (2023) 1–16.

[4] Z. Jiang, F.F. Xu, J. Araki, G. Neubig, How can we know what language models know? Trans. Assoc. Comput. Linguist. 8 (2020) 423–438.

[5] Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., and Dong, Z.: A survey of large language models, arXiv preprint arXiv:2303.18223, 2023.

[6] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang, A survey on evaluation of large language models, ACM Trans. Intell. Syst. Technol. 15 (3) (2024) 1–45.

[7] Mao, R., Chen, G., Zhang, X., Guerin, F., and Cambria, E.: GPTEval: A survey on assessments of ChatGPT and GPT-4, arXiv preprint arXiv:2308.12488, 2023.

[8] Peng, B., Li, C., He, P., Galley, M., and Gao, J.: Instruction tuning with gpt-4, arXiv preprint arXiv:2304.03277, 2023.

[9] A. Turchin, S. Masharsky, M. Zitnik, Comparison of BERT implementations for natural language processing of narrative medical documents, Inform. Med. Unlocked 36 (2023) 101139.

[10] Aftan, S., and Shah, H.: A Survey on BERT and Its Applications, in Editor (Ed.)^(Eds.): Book A Survey on BERT and Its Applications (IEEE, 2023, edn.), pp. 161-166.

[11] Tunstall, L., Von Werra, L., and Wolf, T.: Natural language processing with transformers (" OReilly Media, Inc.", 2022. 2022).

[12] K.M. Jablonka, Q. Ai, A. Al-Feghali, S. Badhwar, J.D. Bocarsly, A.M. Bran, S. Bringuier, L.C. Brinson, K. Choudhary, D. Circi, 14 examples of how LLMs can transform materials science and chemistry: a reflection on a large language model hackathon, Digit. Discov. (2023).

[13] T. Xie, Y. Wa, W. Huang, Y. Zhou, Y. Liu, Q. Linghu, S. Wang, C. Kit, C. Grazian, B. Hoex, Large language models as master key: unlocking the secrets of materials science with GPT, arXiv:2304.02213,, arXiv Prepr. (2023).

[14] J. Schrier, A.J. Norquist, T. Buonassisi, J. Brgoch, In pursuit of the exceptional: research directions for machine learning in chemical and materials science, J. Am. Chem. Soc. (2023).

[15] X. Bai, Y. Xie, X. Zhang, H. Han, J.-R. Li, Evaluation of open-source large language models for metal–organic frameworks research, J. Chem. Inf. Model. (2024).

[16] R.K. Luu, M.J. Buehler, BioinspiredLLM: conversational large language model for the mechanics of biological and bio-inspired materials, Adv. Sci. 11 (10) (2024) 2306724.

[17] J. Choi, B. Lee, Accelerating materials language processing with large language models, Commun. Mater. 5 (1) (2024) 13.

[18] Choi, J., and Lee, B.: Accelerated materials language processing enabled by GPT, arXiv preprint arXiv:2308.09354, 2023.

[19] C.M. Castro Nascimento, A.S. Pimentel, Do large language models understand chemistry? A conversation with ChatGPT, J. Chem. Inf. Model. 63 (6) (2023) 1649–1655.

[20] A.D. White, The future of chemistry is language, Nat. Rev. Chem. (2023) 1–2.

[21] Bran, A.M., Cox, S., White, A.D., and Schwaller, P.: ChemCrow: Augmenting large-language models with chemistry tools, arXiv preprint arXiv:2304.05376, 2023.

[22] N. Sitapure, J.S.-I. Kwon, Exploring the potential of time-series transformers for process modeling and control in chemical systems: an inevitable paradigm shift, Chem. Eng. Res. Des. 194 (2023) 461–477.

[23] Frey, N., Soklaski, R., Axelrod, S., Samsi, S., Gomez-Bombarelli, R., Coley, C., and Gadepally, V.: Neural scaling of deep chemical models, 2022.

[24] Wang, Y., Zhao, Y., and Petzold, L.: Are Large Language Models Ready for Healthcare? A Comparative Study on Clinical Language Understanding, arXiv preprint arXiv:2304.05368, 2023.

[25] Jang, D., and Kim, C.-E.: Exploring the Potential of Large Language models in Traditional Korean Medicine: A Foundation Model Approach to Culturally-Adapted Healthcare, arXiv preprint arXiv:2303.17807, 2023.

[26] A.J. Thirunavukarasu, D.S.J. Ting, K. Elangovan, L. Gutierrez, T.F. Tan, D.S. W. Ting, Large language models in medicine, Nat. Med. (2023) 1–11.

[27] Prasad, N., Boughanem, M., and Dkaki, T.: Effect of hierarchical domain-specific language models and attention in the classification of decisions for legal cases, in Editor (Ed.)^(Eds.): 'Book Effect of hierarchical domain-specific language models and attention in the classification of decisions for legal cases' (2022, edn.), pp. 4-7.

[28] Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., and Mann, G.: Bloomberggpt: A large language model for finance, arXiv preprint arXiv:2303.17564, 2023.

[29] Zhang, L., Cai, W., Liu, Z., Yang, Z., Dai, W., Liao, Y., Qin, Q., Li, Y., Liu, X., and Liu, Z.: FinEval: A Chinese Financial Domain Knowledge Evaluation Benchmark for Large Language Models, arXiv preprint arXiv:2308.09975, 2023.

[30] T. Gupta, M. Zaki, N.A. Krishnan, Mausam, MatSciBERT: a materials domain language model for text mining and information extraction, NPJ Comput. Mater. 8 (1) (2022) 102.

[31] S. Pal, M. Bhattacharya, S.-S. Lee, C. Chakraborty, A domain-specific next-generation large language model (LLM) or ChatGPT is required for biomedical engineering and research, Ann. Biomed. Eng. (2023) 1–4.

[32] Wang, Z., Yang, F., Zhao, P., Wang, L., Zhang, J., Garg, M., Lin, Q., and Zhang, D.: Empower Large Language Model to Perform Better on Industrial Domain-Specific Question Answering, arXiv preprint arXiv:2305.11541, 2023.

[33] E.O. Pyzer-Knapp, J.W. Pitera, P.W. Staar, S. Takeda, T. Laino, D.P. Sanders, J. Sexton, J.R. Smith, A. Curioni, Accelerating materials discovery using artificial intelligence, high performance computing and robotics, NPJ Comput. Mater. 8 (1) (2022) 84.

[34] C. Gao, X. Min, M. Fang, T. Tao, X. Zheng, Y. Liu, X. Wu, Z. Huang, Innovative materials science via machine learning, Adv. Funct. Mater. 32 (1) (2022) 2108044.

[35] S.G. Higgins, A.A. Nogiwa-Valdez, M.M. Stevens, Considerations for implementing electronic laboratory notebooks in an academic research environment, Nat. Protoc. 17 (2) (2022) 179–189.

[36] M. Schröder, S. Staehlke, P. Groth, J.B. Nebe, S. Spors, F. Krüger, Structure-based knowledge acquisition from electronic lab notebooks for research data provenance documentation, J. Biomed. Semant. 13 (1) (2022) 1–22.

[37] S. Herres-Pawlis, F. Bach, I.J. Bruno, S.J. Chalk, N. Jung, J.C. Liermann, L. R. McEwen, S. Neumann, C. Steinbeck, M. Razum, Minimum information standards in chemistry: a call for better research data management practices, Angew. Chem. Int. Ed. 61 (51) (2022) e202203038.

[38] C. Woo, J. Yoo, Exploring the determinants of blockchain acceptance for research data management, J. Comput. Inf. Syst. 63 (1) (2023) 216–227.

[39] N. Brandt, L. Griem, C. Herrmann, E. Schoof, G. Tosato, Y. Zhao, P. Zschumme, M. Selzer, Kadi4Mat: a research data infrastructure for materials science, Data Sci. J. 20 (2021), pp. 8-8.

[40] N. CARP, A. Minges, M. Piel, eLabFTW: an open source laboratory notebook for research labs, J. Open Source Softw. 2 (12) (2017) 146.

[41] N.H. Goddard, R. Macneil, J. Ritchie, eCAT: online electronic lab notebook for scientific research, Autom. Exp. 1 (1) (2009) 1–7.

[42] Carpi, N.: eLabFTW Homepage, 2013.

[43] Corley, C.D., and Mihalcea, R.: Measuring the semantic similarity of texts, in Editor (Ed.)^(Eds.): Book Measuring the semantic similarity of texts (2005, edn.), pp. 13-18.

[44] L.M. Ghiringhelli, C. Baldauf, T. Bereau, S. Brockhauser, C. Carbogno, J. Chamanara, S. Cozzini, S. Curtarolo, C. Draxl, S. Dwaraknath, Shared metadata for data-centric materials science, Sci. Data 10 (1) (2023) 626.

[45] M. Scheidgen, L. Himanen, A.N. Ladines, D. Sikter, M. Nakhaee, Á. Fekete, T. Chang, A. Golparvar, J.A. Márquez, S. Brockhauser, NOMAD: a distributed web-based platform for managing materials science research data, J. Open Source Softw. 8 (90) (2023) 5388.

[46] M. Jalali, M. Mail, R. Aversa, C. Kübel, MSLE: an ontology for materials science laboratory equipment–Large-scale devices for materials characterization, Mater. Today Commun. 35 (2023) 105532.